

```
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WordPerfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/04018
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/221,580
;; FILING DATE: 01-APR-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/221,583
;; FILING DATE: 01-APR-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/221,581
;; FILING DATE: 01-APR-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: DeLuca, Mark
;; REGISTRATION NUMBER: 33,229
;; REFERENCE/DOCKET NUMBER: CCOR-0232
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 568-3100
;; TELEFAX: (215) 568-3439
;; INFORMATION FOR SEQ ID NO: 58:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 15 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; PCT-US95-04018-58

Query Match 0.9%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 QPTPEPS 358
Db 5 QPTPEPS 11

RESULT 23
PCT-US95-04018-59
; Sequence 59, Application PC/TUS9504018
; GENERAL INFORMATION:
; APPLICANT: Heavner, George A.
; APPLICANT: Kruszynski, Marian
; APPLICANT: Mervic, Miljenko
; APPLICANT: Weber, Robert W.
; TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: Norris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04018
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,580
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:
```

```
;;
;; APPLICATION NUMBER: US 08/221,583
;; FILING DATE: 01-APR-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/221,581
;; FILING DATE: 01-APR-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: DeLuca, Mark
;; REGISTRATION NUMBER: 33,229
;; REFERENCE/DOCKET NUMBER: CCOR-0232
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 568-3100
;; TELEFAX: (215) 568-3439
;; INFORMATION FOR SEQ ID NO: 59:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 15 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; PCT-US95-04018-59

Query Match 0.9%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 QPTPEPS 358
Db 2 QPTPEPS 8

RESULT 24
US-09-020-116-4
; Sequence 4, Application US/09020116
; Patent No. 6084063
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS
; TITLE OF INVENTION: OF ALLERGIC REACTIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: SUITE 1200, 127 PEACHTREE STREET
; CITY: ATLANTA
; STATE: GA
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLER, MARY L
; REGISTRATION NUMBER: 39,303
; REFERENCE/DOCKET NUMBER: 14014.0285
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-221-583-58

Query Match 0.9%; Score 7; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QPTPEPS 358
Db 5 QPTPEPS 11

RESULT 20
US-08-221-583-59
Sequence 59, Application US/08221583
Patent No. 5486595
GENERAL INFORMATION:
APPLICANT: Heavner, George A.
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19403

COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25:mdctcMod.
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/221,583
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: DeLuca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: CCOR-0185
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-221-583-59

Query Match 0.9%; Score 7; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QPTPEPS 358
Db 2 QPTPEPS 8

RESULT 21
PCT-US95-04018-57
Sequence 57, Application PC/TUS9504018
GENERAL INFORMATION:
APPLICANT: Heavner, George A.

APPLICANT: Kruszynski, Marian
APPLICANT: Mervic, Miljenko
APPLICANT: Weber, Robert W.
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19403

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04018
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,580
FILING DATE: 01-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,583
FILING DATE: 01-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,581
FILING DATE: 01-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: DeLuca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: CCOR-0232
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-04018-57

Query Match 0.9%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QPTPEPS 358
Db 8 QPTPEPS 14

RESULT 22
PCT-US95-04018-58
Sequence 58, Application PC/TUS9504018
GENERAL INFORMATION:
APPLICANT: Heavner, George A.
APPLICANT: Kruszynski, Marian
APPLICANT: Mervic, Miljenko
APPLICANT: Weber, Robert W.
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19403

; LENGTH: 490
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5116

Query Match 1.0%; Score 8; DB 4; Length 490;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 NDGAVALA 157
Db 72 NDGAVALA 79

RESULT 17
US-08-999-774A-10
; Sequence 10, Application US/08999774A
; Patent No. 6274312
; GENERAL INFORMATION:
; APPLICANT: Gish, Kurt C.
; APPLICANT: Seghezzi, Wolfgang
; APPLICANT: Shanahan, Frances
; APPLICANT: Lees, Emma M.
; APPLICANT: McCleanhan, Terrill K.
; TITLE OF INVENTION: Intracellular Regulatory Molecules;
; TITLE OF INVENTION: Related Reagents
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,774A
; FILING DATE: 10-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,818
; FILING DATE: 11-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1200
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-999-774A-10

Query Match 1.0%; Score 8; DB 3; Length 503;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 357 PSPGPQA 364
Db 244 PSPGPQA 251

RESULT 18
US-08-221-583-57
; Sequence 57, Application US/08221583

; Patent No. 5486595
; GENERAL INFORMATION:
; APPLICANT: Heavner, George A.
; TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25:mdctcMod.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,583
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: CCOR-0185
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-221-583-57

Query Match 0.9%; Score 7; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 QPTPEPS 358
Db 8 QPTPEPS 14

RESULT 19
US-08-221-583-58
; Sequence 58, Application US/08221583
; Patent No. 5486595
; GENERAL INFORMATION:
; APPLICANT: Heavner, George A.
; TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25:mdctcMod.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,583
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: CCOR-0185

Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TNSNNSNT 257
|||||
Db 178 TNSNNSNT 186

RESULT 14

US-08-569-166-34
; Sequence 34, Application US/08569166
; Patent No. 5830722
; GENERAL INFORMATION:
; APPLICANT: NICOLAS, LUC
; APPLICANT: CHARLES, JEAN-FRANCOIS
; APPLICANT: DECLEUSE, ARMELE
; APPLICANT: BARLOY, FREDERIQUE
; TITLE OF INVENTION: CLOSTRIDIUM BIFERMENTANS DNA FRAGMENT
; TITLE OF INVENTION: BEARING GENES CODING FOR PROTEINS LINKED TO AN
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,166
; FILING DATE: 05-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR94/00768
; FILING DATE: 24-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93/07795
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-106-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 110 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-166-34

Query Match 1.0%; Score 8; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 TNSNNSNT 257
|||||
Db 67 TNSNNSNT 74

RESULT 15

US-09-107-532A-5477
; Sequence 5477, Application US/09107532A
; Patent No. 6583275

GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354

; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5477:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium

; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...332
; SEQUENCE DESCRIPTION: SEQ ID NO: 5477:
US-09-107-532A-5477

Query Match 1.0%; Score 8; DB 4; Length 332;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 EEINROKQ 137
|||||
Db 137 EEINROKQ 144

RESULT 16

US-09-134-001C-5116
; Sequence 5116, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354

; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,001C
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/064,964
; FILING DATE: 1997-11-08
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-007

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5477:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...332
; SEQUENCE DESCRIPTION: SEQ ID NO: 5477:
US-09-107-532A-5477

US-09-107-532A-5477


```
Query Match          4.0%; Score 32; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 5.2e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 AEQIVIKITDGYVTSHGDHYHYNGKVPYDA 77
      |||||
Db 67 AEQIVIKITDGYVTSHGDHYHYNGKVPYDA 98

RESULT 11
US-09-328-352-6143
; Sequence 6143, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6143
; LENGTH: 795
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6143

Query Match          1.1%; Score 9; DB 4; Length 795;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 201 SASBLAAAE 209
      |||||
Db 428 SASBLAAAE 436

RESULT 12
US-08-188-582-2
; Sequence 2, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 36
; ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 28-JAN-1994
; APPLICATION NUMBER: US/08/188,582
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO

Query Match          1.1%; Score 9; DB 1; Length 921;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 249 TNTSNNSNT 257
      |||||
Db 178 TNTSNNSNT 186

RESULT 13
US-08-646-715-2
; Sequence 2, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-646-715-2

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-582-2
```

```
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 763 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-536-784-66

Query Match          6.4%; Score 51; DB 4; Length 763;
Best Local Similarity 100.0%; Pred. No. 1.2e-41;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 QGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNLSASELAARAA 210
DB 159 QGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNLSASELAARAA 209

RESULT 8
US-09-961-083-182
; Sequence 182, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 447 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-536-784-182

Query Match          4.0%; Score 32; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 4.8e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEQIVIKITDQGYVTSBGHDHYHYNGKVPYDA 77
DB 43 AEQIVIKITDQGYVTSBGHDHYHYNGKVPYDA 74

RESULT 10
US-09-468-656A-6
; Sequence 6, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-6
```

```
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-10

Query Match          7.5%; Score 60; DB 4; Length 819;
Best Local Similarity 100.0%; Pred. No. 1.6e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISSELLMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISSELLMKDPNY 111

RESULT 5
US-09-468-656A-4
; Sequence 4, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-4

Query Match          7.5%; Score 60; DB 4; Length 838;
Best Local Similarity 100.0%; Pred. No. 1.7e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISSELLMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISSELLMKDPNY 111

RESULT 6
US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512

; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
```

```
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-66

Query Match          6.4%; Score 51; DB 3; Length 763;
Best Local Similarity 100.0%; Pred. No. 1.2e-41;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 QGRYTTDGYIFNASDIIEDTGDYIVPHGDHYHYIPKNELSAELAAEA 210
Db 159 QGRYTTDGYIFNASDIIEDTGDYIVPHGDHYHYIPKNELSAELAAEA 209

RESULT 7
US-09-536-784-66
; Sequence 66, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
```

| | | | |
|----|-----|--|-----|
| Qy | 1 | SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT | 60 |
| Db | 1 | SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT | 60 |
| Qy | 61 | SHGDHYHYNGKVPYDAIISBELLMKDPNYKKBEDI VNEVKGGYVIKVDGKYVYVLKDA | 120 |
| Db | 61 | SHGDHYHYNGKVPYDAIISBELLMKDPNYKKBEDI VNEVKGGYVIKVDGKYVYVLKDA | 120 |
| Qy | 121 | AHADNVRTKEEINRQKQHSQHRGGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT | 180 |
| Db | 121 | AHADNVRTKEEINRQKQHSQHRGGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT | 180 |
| Qy | 181 | GDAYIVPHGDHYHYIPKNELASAEIAAEAFISGRGNLSNGTYRQNSDNTSRTNWVPS | 240 |
| Db | 181 | GDAYIVPHGDHYHYIPKNELASAEIAAEAFISGRGNLSNGTYRQNSDNTSRTNWVPS | 240 |
| Qy | 241 | VSNGTNTNTNNSNTNNSQASQNSDIDSLLKQLYKLPFSQRHVESDGLVFPDPAQITST | 300 |
| Db | 241 | VSNGTNTNTNNSNTNNSQASQNSDIDSLLKQLYKLPFSQRHVESDGLVFPDPAQITST | 300 |
| Qy | 301 | ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHNWVPSDEPQSPSPOTPEPSG | 360 |
| Db | 301 | ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHNWVPSDEPQSPSPOTPEPSG | 360 |
| Qy | 361 | PQAPAPNLKIDNSNLSVLVSQLVKVGEGYVFEKGISRYVFAKOLPSETVKNLESKLSKQBS | 420 |
| Db | 361 | PQAPAPNLKIDNSNLSVLVSQLVKVGEGYVFEKGISRYVFAKOLPSETVKNLESKLSKQBS | 420 |
| Qy | 421 | VSHTLTAKENVAPRQDFYDQKAYNLLTEAHKALFXNKGKNSDFOALDKLLERLNDESTN | 480 |
| Db | 421 | VSHTLTAKENVAPRQDFYDQKAYNLLTEAHKALFXNKGKNSDFOALDKLLERLNDESTN | 480 |
| Qy | 481 | KEKLVDDLLAPLATPTTPERLGKPNSTQYETDEVRIAQALADKYTTSDGYIFDEHDIISD | 540 |
| Db | 481 | KEKLVDDLLAPLATPTTPERLGKPNSTQYETDEVRIAQALADKYTTSDGYIFDEHDIISD | 540 |
| Qy | 541 | EGDAYVTPHMGHSWIGKDSLSDKBKAAQAYTKEKGILPPSPDADVKNPAGTGSAAAY | 600 |
| Db | 541 | EGDAYVTPHMGHSWIGKDSLSDKBKAAQAYTKEKGILPPSPDADVKNPAGTGSAAAY | 600 |
| Qy | 601 | NRVKGEXRIPLVRLPYMVEHTVEVKNGNLIIPH KDHYHNIKEAFWDDHTYKAPNGYTTLED | 660 |
| Db | 601 | NRVKGEXRIPLVRLPYMVEHTVEVKNGNLIIPH KDHYHNIKEAFWDDHTYKAPNGYTTLED | 660 |
| Qy | 661 | LFATIKYVVEHPDRPHSNDGWGNASHVLGKQHSDEDPKNKFKADEBEVESTPAPEVP | 720 |
| Db | 661 | LFATIKYVVEHPDRPHSNDGWGNASHVLGKQHSDEDPKNKFKADEBEVESTPAPEVP | 720 |
| Qy | 721 | QVETEKVEAQLKEAEVLLAKVYDSSLSKANATETLAGLRNLLTQIIMDNNSIMAEAEKLLA | 780 |
| Db | 721 | QVETEKVEAQLKEAEVLLAKVYDSSLSKANATETLAGLRNLLTQIIMDNNSIMAEAEKLLA | 780 |
| Qy | 781 | LLKGSNPSSVSKEKIN | 796 |
| Db | 781 | LLKGSNPSSVSKEKIN | 796 |

RESULT 3

```

RESUL 3
US-09-468-656A-8
; Sequence 8, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositi
; TITLE OF INVENTION: Pneumoniae Gro
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468-656A-8
; CURRENT FILING DATE: 1998-12-02
; PRIOR APPLICATION NUMBER: 60/113,041
; PRIOR FILING DATE: 1998-12-21

```

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-8

```

Query Match 87.3%; Score 695; DB 4; Length 819;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795: Conservative 0; Mismatches 1; Indels

| | | | | | | |
|----|-----|----------------------|----------------------|-------------------------------|--------------|----|
| Qy | 1 | SYELGHIQARTYKKNRVSII | IGDKQATQKTENI | TPDEVSKREGINAEOI | IVIKITDOGYVT | 60 |
| Db | 21 | SYELGHIQARTYKKNRVSII | IGDKQATQKTENI | TPDEVSKREGINAEOI | IVIKITDOGYVT | 80 |
| Qy | 61 | SHGDHYHYNGKVPYDAII | ISELLMKDPNYKLKDEI | VNEVKGYYIKVDGYVYVLKDA | 120 | |
| Db | 81 | SHGDHYHYNGKVPYDAII | ISELLMKDPNYKLKDEI | VNEVKGYYIKVDGYVYVLKDA | 140 | |
| Qy | 121 | AHADNVRTKEENRQKEHS | QHRREGTTPNDGAVALAR | SQGRYTTDDGYIFNASDIEUT | 180 | |
| Db | 141 | AHADNVRTKEENRQKEHS | QHRREGTTPNDGAVALAR | SQGRYTTDDGYIFNASDIEUT | 200 | |
| Qy | 181 | GDAYIVPHGDHYHYPKNLS | ASELAEEAFISGRGNLS | NSRTRYRQNSDNTSRTNWVPS | 240 | |
| Db | 201 | GDAYIVPHGDHYHYPKNLS | ASELAEEAFISGRGNLS | NSRTRYRQNSDNTSRTNWVPS | 260 | |
| Qy | 241 | VSNPGTNTNTNNSNTNS | QAQSNDIDSLAQLYKLPLS | QRHVSDGLVFPAQITSRT | 300 | |
| Db | 261 | VSNPGTNTNTNNSNTNS | QAQSNDIDSLAQLYKLPLS | QRHVSDGLVFPAQITSRT | 320 | |
| Qy | 301 | ARGVAVPHGDHYHYPYOMSE | LEERIARIIPLYRSNHWVPS | DRPEQSPQTPPEPSPG | 360 | |
| Db | 321 | ARGVAVPHGDHYHYPYOMSE | LEERIARIIPLYRSNHWVPS | DRPEQSPQTPPEPSPG | 380 | |
| Qy | 361 | POPAPNLKIDSNSLSV | SQLVRKVGEGYVEEKGISRY | VEFAKDLPSRTVKNLESKLSQBS | 420 | |
| Db | 381 | POPAPNLKIDSNSLSV | SQLVRKVGEGYVEEKGISRY | VEFAKDLPSRTVKNLESKLSQBS | 440 | |
| Qy | 421 | VSHTLTAKKENVAPRQOE | FYDKAYNLLTEAHKALF | XNKGNSDFQALDKLLERLNDESTN | 480 | |
| Db | 441 | VSHTLTAKKENVAPRQOE | FYDKAYNLLTEAHKALF | XNKGNSDFQALDKLLERLNDESTN | 500 | |
| Qy | 481 | KEKLVDDLLAFIAPITH | PERLKGPNQSI EYTED | EVRIAOLAKYTTSDGYIFDEHDI | 540 | |
| Db | 501 | KEKLVDDLLAFIAPITH | PERLKGPNQSI EYTED | EVRIAOLAKYTTSDGYIFDEHDI | 560 | |
| Qy | 541 | EGDAYYTPMHGSHWIGK | LSDKBKVAAQAYTKEKGIL | PPSPDADVKANPTGDSAAAIY | 600 | |
| Db | 561 | EGDAYYTPMHGSHWIGK | LSDKBKVAAQAYTKEKGIL | PPSPDADVKANPTGDSAAAIY | 620 | |
| Qy | 601 | NRVKGKRIPLVRLPYM | VEHTVEVKNGNLIIPH | KDHYHNIKFAWFDHDTYKAPNGYTL | 660 | |
| Db | 621 | NRVKGKRIPLVRLPYM | VEHTVEVKNGNLIIPH | KDHYHNIKFAWFDHDTYKAPNGYTL | 680 | |
| Qy | 661 | LFATIKYVEHPDERPHS | NDGWNASEHVLGKDH | SEDPKNFKADEEVEETPAEPYV | 720 | |
| Db | 681 | LFATIKYVEHPDERPHS | NDGWNASEHVLGKDH | SEDPKNFKADEEVEETPAEPYV | 740 | |
| Qy | 721 | QVETEKVEAQLKEAEV | LLAKVTDSSILKANAT | ETIAGLRNNLTLQIMDNNSIMAEAEKLLA | 780 | |
| Db | 741 | QVETEKVEAQLKEAEV | LLAKVTDSSILKANAT | ETIAGLRNNLTLQIMDNNSIMAEAEKLLA | 800 | |
| Qy | 781 | LLKGSNPSSVSKEKIN | 796 | | | |
| Db | 801 | LLKGSNPSSVSKEKIN | 816 | | | |

RESULT 4
US-09-468-656A-10
; Sequence 10, Application US/09468656A
; Patent No. 6582706

```

; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-56

Query Match 99.9%; Score 795; DB 3; Length 796;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYELGLQARTVKNNRVSIDGQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 1 SYELGLQARTVKNNRVSIDGQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Qy 61 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLDKEDIIVNEKGGYVIKVDGKYVYLKDA 120
Db 61 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLDKEDIIVNEKGGYVIKVDGKYVYLKDA 120
Qy 121 AHADNVRTKEINRQKQHSQHRGGTTRNDGAVALARQSGRYTTDDGYIFNASDIIEDT 180
Db 121 AHADNVRTKEINRQKQHSQHRGGTTRNDGAVALARQSGRYTTDDGYIFNASDIIEDT 180
Qy 181 GDVIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Db 181 GDVIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Qy 241 VSNPQTNTNTSNNTNSQASQNSQNDISLLKQLYKPLSRHVESDGLVDPDPAQITSR 300
Db 241 VSNPQTNTNTSNNTNSQASQNSQNDISLLKQLYKPLSRHVESDGLVDPDPAQITSR 300
Qy 301 ARGVAVPHGDHYHYIPYSQMSLEBERIARIIPLYRSNHHVDPDRPEOPSQPTPEPSPG 360
Db 301 ARGVAVPHGDHYHYIPYSQMSLEBERIARIIPLYRSNHHVDPDRPEOPSQPTPEPSPG 360
Qy 361 PQAPNLKIDSNLSVLSQVRKVGEGYVFEKSGISRYVFAKDLFSETVKNLESKLSKQES 420
Db 361 PQAPNLKIDSNLSVLSQVRKVGEGYVFEKSGISRYVFAKDLFSETVKNLESKLSKQES 420
Qy 421 VSHTLTAKENVAPRQDFYKAYNLLTEAHKALFYXNKGKNSDFQALDKLLERLNDESTN 480
Db 421 VSHTLTAKENVAPRQDFYKAYNLLTEAHKALFYXNKGKNSDFQALDKLLERLNDESTN 480

```

```

Qy 481 KEKLVDDLLAFAPITHPERLGPNSQIEYTEDEVRIQAQADKYTTSDGYIFDEHDIISD 540
Db 481 KEKLVDDLLAFAPITHPERLGPNSQIEYTEDEVRIQAQADKYTTSDGYIFDEHDIISD 540
Qy 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQYTKKGIPLPPSPDADVKANPTGDSAAAIY 600
Db 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQYTKKGIPLPPSPDADVKANPTGDSAAAIY 600
Qy 601 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPKHQHYHNIKFAWFDHHTYKAPNGYTLED 660
Db 601 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPKHQHYHNIKFAWFDHHTYKAPNGYTLED 660
Qy 661 LFATIKYVVEHPDERPHSNDGNGNASEHVLGKKHSDSDPNKFKADBEPEVEETPAEPEVP 720
Db 661 LFATIKYVVEHPDERPHSNDGNGNASEHVLGKKHSDSDPNKFKADBEPEVEETPAEPEVP 720
Qy 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780
Db 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780
Qy 781 LLKGSNPSSVSKEKIN 796
Db 781 LLKGSNPSSVSKEKIN 796

RESULT 2
US-09-536-784-56
; Sequence 56, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-536-784-56

```

Query Match 99.9%; Score 795; DB 4; Length 796;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:31:07 ; Search time 24 Seconds
(without alignments)
1712.261 Million cell updates/sec

Title: US-09-765-271-56
Perfect score: 796
Sequence: 1 SYELGLYQARTVKENRVS.....KLLALLKGSNPSSVSKEKIN 796

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : Issued Patents_AA:*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 795 | 99.9 | 796 | 3 | US-08-961-083-56 |
| 2 | 795 | 99.9 | 796 | 4 | US-09-536-784-56 |
| 3 | 695 | 87.3 | 819 | 4 | US-09-468-656A-8 |
| 4 | 60 | 7.5 | 819 | 4 | US-09-468-656A-10 |
| 5 | 60 | 7.5 | 838 | 4 | US-09-468-656A-4 |
| 6 | 51 | 6.4 | 763 | 3 | US-08-961-083-66 |
| 7 | 51 | 6.4 | 763 | 4 | US-09-536-784-66 |
| 8 | 32 | 4.0 | 447 | 3 | US-08-961-083-182 |
| 9 | 32 | 4.0 | 447 | 4 | US-09-536-784-182 |
| 10 | 32 | 4.0 | 484 | 4 | US-09-468-656A-6 |
| 11 | 9 | 1.1 | 795 | 4 | US-09-328-352-6143 |
| 12 | 9 | 1.1 | 921 | 1 | US-08-188-582-2 |
| 13 | 9 | 1.1 | 921 | 1 | US-08-546-715-2 |
| 14 | 8 | 1.0 | 110 | 2 | US-08-569-166-34 |
| 15 | 8 | 1.0 | 332 | 4 | US-09-107-532A-5477 |
| 16 | 8 | 1.0 | 490 | 4 | US-09-134-001C-5116 |
| 17 | 8 | 1.0 | 503 | 3 | US-08-999-774A-10 |
| 18 | 7 | 0.9 | 15 | 1 | US-08-221-583-57 |
| 19 | 7 | 0.9 | 15 | 1 | US-08-221-583-58 |
| 20 | 7 | 0.9 | 15 | 1 | US-08-221-583-59 |
| 21 | 7 | 0.9 | 15 | 5 | PCT-US95-04018-57 |
| 22 | 7 | 0.9 | 15 | 5 | PCT-US95-04018-58 |
| 23 | 7 | 0.9 | 15 | 5 | PCT-US95-04018-59 |
| 24 | 7 | 0.9 | 27 | 3 | US-09-020-116-4 |
| 25 | 7 | 0.9 | 27 | 4 | US-09-608-902-4 |
| 26 | 7 | 0.9 | 45 | 3 | US-09-020-116-2 |
| 27 | 7 | 0.9 | 45 | 4 | US-09-608-902-2 |

ALIGNMENTS

RESULT 1
US-08-961-083-56
; Sequence 56, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:

| | | | | | | |
|----|---|-----|-----|---|----------------------|-------------------|
| 28 | 7 | 0.9 | 50 | 3 | US-09-171-646-1 | Sequence 1, Appli |
| 29 | 7 | 0.9 | 66 | 3 | US-09-020-116-1 | Sequence 1, Appli |
| 30 | 7 | 0.9 | 66 | 4 | US-09-608-902-1 | Sequence 1, Appli |
| 31 | 7 | 0.9 | 76 | 4 | US-09-328-352-6019 | Sequence 6019, Ap |
| 32 | 7 | 0.9 | 117 | 4 | US-09-134-001C-4585 | Sequence 4585, Ap |
| 33 | 7 | 0.9 | 141 | 4 | US-09-543-681A-5683 | Sequence 5683, Ap |
| 34 | 7 | 0.9 | 149 | 4 | US-09-252-991A-28698 | Sequence 28698, A |
| 35 | 7 | 0.9 | 158 | 2 | US-08-917-456-2 | Sequence 2, Appli |
| 36 | 7 | 0.9 | 158 | 3 | US-09-229-804-2 | Sequence 2, Appli |
| 37 | 7 | 0.9 | 163 | 4 | US-09-252-991A-20622 | Sequence 20622, A |
| 38 | 7 | 0.9 | 191 | 4 | US-09-252-991A-30217 | Sequence 30217, A |
| 39 | 7 | 0.9 | 198 | 4 | US-09-613-303-35 | Sequence 35, Appl |
| 40 | 7 | 0.9 | 198 | 4 | US-10-267-311-35 | Sequence 35, Appl |
| 41 | 7 | 0.9 | 201 | 2 | US-08-916-901-3 | Sequence 3, Appli |
| 42 | 7 | 0.9 | 201 | 2 | US-08-916-901-8 | Sequence 3, Appli |
| 43 | 7 | 0.9 | 201 | 4 | US-09-154-602-3 | Sequence 8, Appli |
| 44 | 7 | 0.9 | 201 | 4 | US-09-154-602-8 | Sequence 8, Appli |
| 45 | 7 | 0.9 | 226 | 3 | US-08-908-332-5 | Sequence 5, Appli |
| 46 | 7 | 0.9 | 235 | 4 | US-09-326-394-4 | Sequence 4, Appli |
| 47 | 7 | 0.9 | 235 | 4 | US-09-580-235-2 | Sequence 2, Appli |
| 48 | 7 | 0.9 | 235 | 4 | US-09-580-235-4 | Sequence 4, Appli |
| 49 | 7 | 0.9 | 235 | 4 | US-09-580-235-6 | Sequence 6, Appli |
| 50 | 7 | 0.9 | 235 | 4 | US-09-580-235-8 | Sequence 8, Appli |
| 51 | 7 | 0.9 | 235 | 4 | US-09-580-181-2 | Sequence 2, Appli |
| 52 | 7 | 0.9 | 235 | 4 | US-09-580-181-4 | Sequence 4, Appli |
| 53 | 7 | 0.9 | 235 | 4 | US-09-580-181-6 | Sequence 6, Appli |
| 54 | 7 | 0.9 | 235 | 4 | US-09-580-181-8 | Sequence 8, Appli |
| 55 | 7 | 0.9 | 235 | 4 | US-09-102-530-2 | Sequence 2, Appli |
| 56 | 7 | 0.9 | 235 | 4 | US-09-102-530-4 | Sequence 4, Appli |
| 57 | 7 | 0.9 | 235 | 4 | US-09-102-530-6 | Sequence 6, Appli |
| 58 | 7 | 0.9 | 235 | 4 | US-09-102-530-8 | Sequence 8, Appli |
| 59 | 7 | 0.9 | 235 | 4 | US-09-620-405B-487 | Sequence 487, App |
| 60 | 7 | 0.9 | 235 | 4 | US-09-604-287A-487 | Sequence 487, App |
| 61 | 7 | 0.9 | 235 | 4 | US-09-834-759-487 | Sequence 487, App |
| 62 | 7 | 0.9 | 240 | 4 | US-09-107-532A-6634 | Sequence 6634, Ap |
| 63 | 7 | 0.9 | 245 | 4 | US-09-252-991A-29404 | Sequence 29404, A |
| 64 | 7 | 0.9 | 255 | 4 | US-09-489-039A-13663 | Sequence 13663, A |
| 65 | 7 | 0.9 | 257 | 4 | US-09-579-845-10 | Sequence 10, Appl |
| 66 | 7 | 0.9 | 261 | 1 | US-08-622-353-2 | Sequence 2, Appli |
| 67 | 7 | 0.9 | 261 | 2 | US-08-622-352A-2 | Sequence 2, Appli |
| 68 | 7 | 0.9 | 261 | 3 | US-08-826-390-2 | Sequence 2, Appli |
| 69 | 7 | 0.9 | 298 | 4 | US-09-252-991A-21944 | Sequence 21944, A |
| 70 | 7 | 0.9 | 305 | 4 | US-09-252-991A-17702 | Sequence 17702, A |
| 71 | 7 | 0.9 | 312 | 4 | US-09-252-991A-31029 | Sequence 31029, A |
| 72 | 7 | 0.9 | 325 | 4 | US-09-252-991A-29807 | Sequence 29807, A |
| 73 | 7 | 0.9 | 364 | 4 | US-09-134-000C-4275 | Sequence 4275, Ap |
| 74 | 7 | 0.9 | 381 | 4 | US-09-198-452A-745 | Sequence 745, App |
| 75 | 7 | 0.9 | 383 | 4 | US-09-459-749D-17 | Sequence 17, Appl |
| 76 | 7 | 0.9 | 390 | 3 | US-08-861-774E-82 | Sequence 82, Appl |
| 77 | 7 | 0.9 | 402 | 4 | US-09-721-870-46 | Sequence 46, Appl |
| 78 | 7 | 0.9 | 406 | 4 | US-08-311-731A-12 | Sequence 12, Appl |
| 79 | 7 | 0.9 | 425 | 4 | US-09-721-870-109 | Sequence 109, App |
| 80 | 7 | 0.9 | 434 | 4 | US-09-489-039A-11674 | Sequence 11674, A |
| 81 | 7 | 0.9 | 447 | 4 | US-09-252-991A-22113 | Sequence 22113, A |
| 82 | 7 | 0.9 | 461 | 1 | US-08-385-229-2 | Sequence 2, Appli |
| 83 | 7 | 0.9 | 461 | 3 | US-08-650-000-2 | Sequence 2, Appli |
| 84 | 7 | 0.9 | 461 | 3 | US-09-042-785A-7 | Sequence 7, Appli |
| 85 | 7 | 0.9 | 461 | 3 | US-08-477-347-3 | Sequence 3, Appli |
| 86 | 7 | 0.9 | 461 | 3 | US-09-006-353A-4 | Sequence 4, Appli |
| 87 | 7 | 0.9 | 461 | 3 | US-08-476-862-2 | Sequence 2, Appli |
| 88 | 7 | 0.9 | 461 | 4 | US-09-573-986-4 | Sequence 4, Appli |
| 89 | 7 | 0.9 | 461 | 4 | US-08-406-824A-2 | Sequence 2, Appli |
| 90 | 7 | 0.9 | 461 | 4 | US-09-800-909-2 | Sequence 2, Appli |

Fri Oct 1 15:35:50 2004

us-09-765-271-56.oli.rsp

Page 16

Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 GVAVPHG 309
Db 59 GVAVPHG 65

RESULT 24
MUTT STRAM
ID MUTT STRAM STANDARD; PRT; 154 AA.
AC F32091;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MUTT-like protein (ORF154).
OS Streptomyces ambifaciens.
OG Plasmid pSAM2.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1889;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23877;
RX MEDLINE=95020551; PubMed=7934842;
RA Hagege J., Pernodet J.L., Friedmann A., Guerinneau M.;
RT "Mode and origin of replication of pSAM2, a conjugative integrating
element of Streptomyces ambifaciens.";
RL Mol. Microbiol. 10:799-812(1993).
CC -!- SIMILARITY: Belongs to the NUDIX hydrolase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z19590; CAA79638.1; --
CC HSP; S39873; S39873.
CC PIR; S39873; ITUM.
CC InterPro; IPR000086; NUDIX_hydrolase.
CC Pfam; PF00293; NUDIX; 1.
CC PRINTS; PR00502; NUDIXFAMILY.
CC PROSITE; PS00893; NUDIX; 1.
KW Plasmid; Hydrolase.
FT DOMAIN 48 69 NUDIX BOX.
SQ SEQUENCE 154 AA; 16689 MW; 8EC2EE99D272D656 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 LTPDEVS 39
Db 114 LTPDEVS 120

RESULT 25
GREA STAM
ID GREA STAM STANDARD; PRT; 158 AA.
AC Q99TN9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcription elongation factor greA (Transcript cleavage factor
DE greA).
GN GREA OR SAV1610 OR SA1438 OR MW1560.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

NCBI_TaxID=158878, 158879, 196620;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=MU50 / ATCC 700699, and N315;
RC MEDLINE=21311952; PubMed=11418146;
RX Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kato C.,
RA Sekimizu K., Hirakawa H., Oshima K., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Hayashi H., Hiramatsu K.,
RA Hattori M., Ogasawara N., Hayashi H., Yoshino C., Shiba T.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus".
RL Lancet 357:1225-1240(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA".
RL Lancet 359:1819-1827(2002).
CC -!- FUNCTION: Necessary for efficient RNA polymerase transcription
CC elongation past template-encoded arresting sites. The arresting
CC sites in DNA have the property of trapping a certain fraction of
CC elongating RNA polymerases that pass through, resulting in locked
CC ternary complexes. Cleavage of the nascent transcript by cleavage
CC factors such as greA or greB allows the resumption of elongation
CC from the new 3' terminus. GreA releases sequences of 2 to 3
CC nucleotides (by similarity).
CC -!- SIMILARITY: Belongs to the greA/greB family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF003362; BAB57772.1; --
CC EMBL; AF003134; BAB42702.1; --
CC EMBL; AF004827; BAB95425.1; --
CC PIR; A89943; A89943.
CC HSSP; P21346; 1GRJ.
CC SWISS-2DPAGE; Q99TN9; STAAH.
CC HAMAP; MF 00105; 1.
CC InterPro; IPR006359; GreA.
CC InterPro; IPR001437; GreA_GreB.
CC Pfam; PF01272; GreA_GreB; 1.
CC Pfam; PF03449; GreA_GreB; 1.
CC ProDom; PD004918; GreA_GreB; 1.
CC TIGRFAMs; TIGR01462; greA; 1.
CC PROSITE; PS00829; GREAB; 1.
CC PROSITE; PS00830; GREAB_2; 1.
KW Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT DOMAIN 4 76 COILED COIL (POTENTIAL).
SQ SEQUENCE 158 AA; 17743 MW; EC3B0F0E6238A107 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 IIEDTGD 182
Db 77 IIEDTGD 83

Search completed: October 1, 2004, 07:32:52
Job time : 24 secs

FT ACT SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC...) (PARTIAL).
SQ SEQUENCE 124 AA; 13711 MW; 9435EF532420F852 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 785 SNPSSVS 791
Db 15 SNPSSVS 21

RESULT 22
Y082 RICPR STANDARD; PRT; 143 AA.
AC Q9ZE65;

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP082.

GN RP082.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;

[1]

RN SEQUENCE FROM N.A.
RP STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AJ235270; CAAL4552.1; -.

DR PIR; A71717; A71717.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 143 AA; 16939 MW; 57637C7A62F9F10 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 143;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 NRQKEH 139
Db 106 NRQKEH 112

RESULT 23

PTVA ECOLI

AC P32155; P76776; STANDARD; PRT; 148 AA.

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE PTS system, fructose-like-1 IIA component (Phosphotransferase enzyme

DE II, A component) (EC 2.7.1.69).

GN FRVA OR B3900.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=93374854; PubMed=836120;
RA Moralejo P., Egan S.M., Hidalgo E.F., Aguilar J.;
RT "Sequencing and characterization of a gene cluster encoding the
RT enzymes for L-rhamnose metabolism in Escherichia coli.";
RL J. Bacteriol. 175:5585-5594(1993).

[2]

RN SEQUENCE FROM N.A.

RP STRAIN=K12 / MG1655;

RX MEDLINE=93347969; PubMed=8346018;

RA Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;

RT "Analysis of the Escherichia coli genome. III. DNA sequence of the

RT region from 87.2 to 89.2 minutes.";

RL Nucleic Acids Res. 21:3391-3398(1993).

[3]

RN REVISIONS TO 81 AND 104-108.

RP STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

[4]

DISCUSSION OF SEQUENCE.

RX MEDLINE=94290319; PubMed=8019415;

RA Reizer J., Michotey V., Reizer A., Saier M.H. Jr.;

RT "Novel phosphotransferase system genes revealed by bacterial genome

RT analysis: unique, putative fructose- and glucoside-specific

RT systems.";

RL Protein Sci. 3:440-450(1994).

CC -!- FUNCTION: This is a component of the phosphoenolpyruvate-dependent
CC sugar phosphotransferase system (PTS), a major carbohydrate active
CC -transport system. The IID domains contain the sugar binding site
CC and the transmembrane channel; the IIA domain contains the primary
CC phosphorylation site (the donor is phospho-HPr); IIA transfers its
CC phosphoryl group to the IIB domain which finally transfers it to
CC the sugar.

CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein

CC histidine + sugar phosphate.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -!- SIMILARITY: Contains 1 PTS EIIA domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X60472; CA43004.1; -.

DR EMBL; L19201; AAB03033.2; -.

DR EMBL; AE000465; AAC76882.1; -.

DR PIR; D48649; D48649.

DR EcoGene; EG11864; frva.

DR InterPro; IPR002178; PTS_EIIA_2.

DR InterPro; IPR004715; PTSIIA fruc.

DR Pfam; PF00359; PTS_EIIA_2; 1.

DR ProDom; PD001699; PTS_EIIA_2; 1.

DR TIGRPFAMs; TIGR00848; fruA; 1.

DR PROSITE; PS00372; PTS_EIIA_2; 1.

KW Phosphotransferase system; Sugar transport; Transferase;

KW Phosphorylation; Complete proteome.

FT MOD RES 64 64 PHOSPHORYLATION (BY SIMILARITY).

FT CONFLICT 81 81 A->G (IN REF. 2).

FT CONFLICT 104 107 QSGE -> KXZ (IN REF. 2).

SQ SEQUENCE 148 AA; 16093 MW; 04AE87B9084BA1C6 CRC64;

Query Match

0.9%; Score 7; DB 1; Length 148;

DR EMBL; D90778; BAA15013.1; ALT INIT.
 DR EMBL; D90779; BAA15019.1; ALT INIT.
 DR EMBL; D90838; BAA15822.1; ALT INIT.
 DR EMBL; D90850; BAA16005.1; ALT INIT.
 DR EMBL; D90851; BAA16013.1; ALT INIT.
 DR EMBL; D90852; BAA16036.1; ALT INIT.
 DR EcoGene; EG40003; insc.
 DR InterPro; IPR002514; Transposase 8.
 DR Pfam; PF01527; Transposase 8; 1.
 KW Transposable element; Transposition; DNA-binding; DNA recombination;
 KW Complete proteome.
 FT VARIANT 34 34 L -> F (IN B1997).
 SQ SEQUENCE 121 AA; 13452 MW; 59431E5C452E067A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 121;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 202 ASELAAA 208
 DB 70 ASELAAA 76

RESULT 20
 INSC SHIFL STANDARD; PRT; 121 AA.
 AC P59444;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transposase insc for insertion element IS2.
 GN (INSC1 OR SF0245) AND (INSC2 OR SF0879) AND (INSC3 OR SF0933) AND
 GN (INSC4 OR SF0960) AND (INSC5 OR SF1054) AND (INSC6 OR SF1165) AND
 GN (INSC7 OR SF1343) AND (INSC8 OR SF1463) AND (INSC9 OR SF1587) AND
 GN (INSC10 OR SF2011) AND (INSC11 OR SF2615) AND
 GN (INSC12 OR SF2694) AND (INSC13 OR SF2873) AND
 GN (INSC14 OR SF2984) AND (INSC15 OR SF3431) AND
 GN (INSC16 OR SF3512) AND (INSC17 OR SF3805) AND
 GN (INSC18 OR SF3873) AND (INSC19 OR SF3988) AND
 GN (INSC20 OR SF4097) AND (INSC21 OR SF4185).
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OC NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 CC -!- FUNCTION: Involved in the transposition of the insertion
 CC sequence IS2 (By similarity).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AE015060; AAN41906.1; ALT INIT.
 DR EMBL; AE015116; AAN42511.1; ALT INIT.
 DR EMBL; AE015123; AAN42562.1; ALT INIT.
 DR EMBL; AE015125; AAN42588.1; ALT INIT.
 DR EMBL; AE015133; AAN42676.1; ALT INIT.
 DR EMBL; AE015143; AAN42781.1; ALT INIT.

DR EMBL; AE015160; AAN42947.1; ALT INIT.
 DR EMBL; AE015170; AAN43060.1; ALT INIT.
 DR EMBL; AE015180; AAN43173.1; ALT INIT.
 DR EMBL; AE015218; AAN43556.1; ALT INIT.
 DR EMBL; AE015277; AAN44112.1; ALT INIT.
 DR EMBL; AE015284; AAN44187.1; ALT INIT.
 DR EMBL; AE015301; AAN44359.1; ALT INIT.
 DR EMBL; AE015311; AAN44465.1; ALT INIT.
 DR EMBL; AE015351; AAN44892.1; ALT INIT.
 DR EMBL; AE015359; AAN44970.1; ALT INIT.
 DR EMBL; AE015387; AAN45245.1; ALT INIT.
 DR EMBL; AE015394; AAN45310.1; ALT INIT.
 DR EMBL; AE015405; AAN45422.1; ALT INIT.
 DR EMBL; AE015418; AAN45522.1; ALT INIT.
 DR EMBL; AE015427; AAN45606.1; ALT INIT.
 DR InterPro; IPR002514; Transposase 8.
 DR Pfam; PF01527; Transposase 8; 1.
 KW Transposable element; Transposition; DNA-binding; DNA recombination;
 KW Complete proteome.
 SQ SEQUENCE 121 AA; 13452 MW; 59431E5C452E067A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 121;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 ASELAAA 208
 DB 70 ASELAAA 76

RESULT 21
 RNP ANTAM STANDARD; PRT; 124 AA.
 AC P00668;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
 GN RNASE1 OR RNS1.
 OS Antilocapra americana (Pronghorn).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Antilocapridae; Antilocapra.
 OC NCBI_TaxID=9891;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Pancreas;
 RX MEDLINE=80075014; PubMed=513141;
 RA Beintema J.J., Gastra W., Munnikema J.;
 RT "Primary structure of pronghorn pancreatic ribonuclease: close
 RT relationship between giraffe and pronghorn.";
 RL J. Mol. Evol. 13:305-316(1979).
 CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
 CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
 CC with 2',3'-cyclic phosphate intermediates.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Pancreas.
 CC -!- SIMILARITY: Belongs to the pancreatic ribonuclease family.
 DR PIR; A00813; NRPRH.
 DR HSP; P00656; LRBG.
 DR InterPro; IPR001427; RNaseA.
 DR Pfam; PF00074; rnaseA; 1.
 DR PRINTS; PR00794; RIBONUCLEASE.
 DR ProDom; PD000535; RNaseA; 1.
 DR SMART; SM00092; RNase_Pc; 1.
 DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
 DR Hydrolase; Nuclease; Endonuclease; Glycoprotein.
 FT DISULFID 26 84 BY SIMILARITY.
 FT DISULFID 40 95 BY SIMILARITY.
 FT DISULFID 58 110 BY SIMILARITY.
 FT DISULFID 65 72 BY SIMILARITY.
 FT ACT_SITE 12 12 BY SIMILARITY.
 FT ACT_SITE 41 41 BY SIMILARITY.

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribonuclease P protein component (EC 3.1.26.5) (RNaseP protein) (RNase
DE P protein) (Protein C5).
GN RNPA OR SE2418.
OS Staphylococcus epidermidis.
CC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qian Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RA "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -|- FUNCTION: RNaseP catalyzes the removal of the 5'-leader sequence
CC from pre-tRNA to produce the mature 5' terminus. It can also cleave
CC other RNA substrates such as 4.5S RNA. The protein component plays
CC an auxiliary but essential role in vivo by binding to the 5'-
CC leader sequence and broadening the substrate specificity of the
CC ribozyme (By similarity).
CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-
CC extra-nucleotide from tRNA precursor.
CC -|- SUBUNIT: Consists of a catalytic RNA component (M1 or rnpB) and a
CC protein subunit (By similarity).
CC -|- SIMILARITY: Belongs to the rnpA family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB016752; AAC06061.1; -;
DR HAMAP; MF 00227; -; 1.
DR InterPro; IPR000100; Ribonuclease P.
DR Pfam; PF00825; Ribonuclease P; 1.
DR ProDom; PD003629; Ribonuclease P; 1.
DR TIGRFAMs; TIGR00188; rnpA; 1.
DR PROSITE; PS00648; RIBONUCLEASE_P; 1.
KW Hydrolase; Nuclease; Endonuclease; tRNA processing; RNA-binding;
KW Complete proteome.
SQ SEQUENCE 115 AA; 13484 MW; 106B2592C8400F18 CRC64;
Query Match 0.9%; Score 7; DB 1; Length 115;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 460 RNSDFQA 466
| | | | |
Db 9 RNSDFQA 15
RESULT 19
INSC ECOLI
ID -INSC ECOLI STANDARD; PRT; 121 AA.
AC P19776; O07989; O08018; O08019; P76357; P77346;
DT 01-FEB-1991 (Rel. 17, Created).
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase insC for insertion element IS2A/D/H/I/K.
GN (INSC1 OR B0360) AND (INSC2 OR B1403) AND (INSC3 OR B1997) AND
GN (INSC4 OR B2861) AND (INSC5 OR B3044) AND (INSC6 OR B4272).
OS Escherichia coli.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88137965; PubMed=2830172;
RA Ronecker H.-J., Rak B.;
RT "Genetic organization of insertion element IS2 based on a revised
RT nucleotide sequence.";
RL Gene 59:291-296(1987).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [5]
RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -|- FUNCTION: Involved in the transposition of the insertion sequence
CC IS2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V00279; CAA23542.1; -;
DR EMBL; U14003; AAA97168.1; ALT_INIT.
DR EMBL; U28377; AAA69212.1; ALT_INIT.
DR EMBL; U28375; AAA83043.1; ALT_INIT.
DR EMBL; AE000143; AAC73463.1; ALT_INIT.
DR EMBL; AE000237; AAC74485.1; ALT_INIT.
DR EMBL; AE000291; AAC75058.1; ALT_INIT.
DR EMBL; AE000369; AAC75900.1; -;
DR EMBL; AE000386; AAC76080.1; -;
DR EMBL; AE000498; AAC77228.1; ALT_INIT.

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z98980; CAB11722.1; --
DR EMBL; AS015167; BAA28750.1; --
DR EMBL; AF192764; AAF19628.1; --
DR EMBL; AF121275; AAD25391.1; --
DR PIR; T38824; T38824.
DR HSP; P35754; 1JHB.
DR GeneDB_SPombe; SPAC4F10.20; --
DR InterPro; IPR002109; Thioredoxin.
DR InterPro; IPR006663; Thioredoxin_dom2.
DR Pfam; PF00462; glutaredoxin; 1.
DR PRINTS; PR00160; GLUTAREDOXIN.
DR PROSITE; PS00195; GLUTAREDOXIN; 1.
KW Redox-active center; Electron transport.
FT DISULFID 25 28 REDOX-ACTIVE (BY SIMILARITY).
FT CONFLICT 52 52 N -> D (IN REF. 3).
SQ SEQUENCE 101 AA; 11261 MW; 30557E19BF33E9BB CRC64;

Query Match 0.9%; Score 7; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 NSDFQAL 467 STANDARD; PRT; 101 AA.
DB 82 NSDFQAL 88

RESULT 16
RS16_UREPA
ID RS16_UREPA STANDARD; PRT; 101 AA.
AC Q9PPS1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S16.
GN RSP OR RPS16 OR UUS68.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;

[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Serovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RA "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum";
RL Nature 407:757-762 (2000).
CC -!- SIMILARITY: Belongs to the S16P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE002155; AAF30982.1; --
DR HSP; P80379; 1ENW.
DR HAMAP; MF_00385; --; 1.
DR InterPro; IPR00307; Ribosomal_S16.
DR Pfam; PF00886; Ribosomal_S16; 1.
DR ProDom; PD003791; Ribosomal_S16; 1.
DR TIGRFAMs; TIGR00002; S16; 1.
DR PROSITE; PS00732; RIBOSOMAL_S16; FALSE_NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 101 AA; 11297 MW; FE9D1E1AB36CBF45 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 101;

Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 PSETVKX 410
DB 66 PSETVKX 72

RESULT 17
YQCC_HABIN STANDARD; PRT; 106 AA.
ID YQCC_HABIN STANDARD; PRT; 106 AA.
AC Q57152; O05061;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein H11436.
GN H11436.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd";
RL Science 269:496-512 (1995).

CC -!- SIMILARITY: TO THE N-TERMINAL OF E.CAROTOVORA EXOENZYME REGULATION
CC REGULON ORF1. THE C-TERMINAL PART IS COLINEAR WITH YQCB.
CC -!- SIMILARITY: STRONG, TO E.COLI YQCC.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U32822; AAC23085.1; --
DR PIR; I64171; I64171.
DR TIGR; H11436; --
DR InterPro; IPR007384; DUF446.
DR Pfam; PF04287; DUF446; 1.
DR PIRSF; PIRSF06257; UCP006257; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 106 AA; 12273 MW; 0955920EBD63228C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 AAEAFLS 213
DB 28 AAEAFLS 34

RESULT 18
RNPA_STAEP STANDARD; PRT; 115 AA.
ID RNPA_STAEP STANDARD; PRT; 115 AA.
AC Q8CMN4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)

Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL J. Bacteriol. 184:5479-5490(2002).
 [7]
 RP SEQUENCE OF 1-15.
 RX MEDLINE=92176646; PubMed=1371791;
 RA Barnes P.F., Mehra V., Rivoire B., Fong S.J., Brennan P.J.,
 RA Voegtline M.S., Minden P., Houghten R.A., Bloom B.R., Modlin R.L.;
 RA "Immunoreactivity of a 10-kDa antigen of Mycobacterium tuberculosis";
 J. Immunol. 148:1835-1840(1992).
 CC -1- FUNCTION: Binds to Cpn60 in the presence of Mg-ATP and suppresses
 CC the ATPase activity of the latter.
 CC -1- SUBUNIT: Heptamer of 7 subunits arranged in a ring (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILILARY: Belongs to the groES chaperonin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X60350; CAA42908.1; -;
 DR EMBL; M25258; AAA25340.1; -;
 DR EMBL; X13739; CAA32003.1; -;
 DR EMBL; Z77165; CAB01005.1; -;
 DR EMBL; AE007158; AAK47865.1; -;
 DR PUR; S01381; BVMYBA
 DR PDB; 1HX5; 08-AUG-01.
 DR PDB; 1P3H; 15-JUL-03.
 DR TIGR; MT3527; -;
 DR TubercuList; RV3418C; -;
 DR HAMAP; MF_00580; -; 1.
 DR InterPro; IPR001476; Chaprnnin_Cpn10.
 DR Pfam; PF00166; cpn10; 1.
 DR PRINTS; PR00297; CHAPERONIN10.
 DR PRODOM; PD000566; Chaprnnin_Cpn10; 1.
 DR PROSITE; PS00681; CHAPERONINS CPN10; 1.
 KW Chaperone; Antigen; Heat shock; Complete proteome; 3D-structure.
 FT INIT MET 0
 SQ SEQUENCE 99 AA; 10673 MW; 1DD128E75CF19AF7 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 99;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 GEEKIPL 611
 |||||
 DB 54 GEEKIPL 60

RESULT 15
 GLRL_SCHPO STANDARD; PRT; 101 AA.
 ID GUR1 SCHPO
 AC Q36032; Q9US58;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutaredoxin 1
 GN GRX1 OR SPAC4F10.20.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 [1]
 RP SEQUENCE FROM N.A.
 RA Kawamukai M.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RA Kim H.-G., Cho Y.-W., Park E.-H., Lim C.-J.;
 RT "Characterization of cDNA encoding thioltransferase (glutaredoxin)
 RT from Schizosaccharomyces pombe";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RA Cho Y.-W., Kim H.-G., Lim C.-J.;
 RT "Isolation and expression of the genomic DNA encoding thioltransferase
 RT (Glutaredoxin) from Schizosaccharomyces pombe";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras K., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: THE DISULFIDE BOND FUNCTIONS AS AN ELECTRON CARRIER IN
 CC THE GLUTATHIONE-DEPENDENT SYNTHESIS OF DEOXYRIBONUCLEOTIDES BY THE
 CC ENZYME RIBONUCLEOTIDE REDUCTASE. IN ADDITION, IT IS ALSO INVOLVED
 CC IN REDUCING SOME DISULFIDES IN A COUPLED SYSTEM WITH GLUTATHIONE
 CC REDUCTASE. THIOLTRANSFERASE CATALYZES CELLULAR THIOL-DISULFIDE
 CC TRANSYDROGENATION REACTIONS. IT TRANSFERS REDUCING EQUIVALENTS
 CC TO CYTOSOLIC PROTEIN AND NONPROTEIN DISULFIDES (BY SIMILARITY).
 CC -1- SIMILARITY: Belongs to the glutaredoxin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ248285; CAB49786.1; -;
DR PIR: A75134; A75134.
DR InterPro: IPR008203; DUF104.
DR Pfam: PF01954; DUF104; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 70 AA; 8009 MW; 4BDE6021B/DFP8032 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 70;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 478 STNKEKL 484 ✓
DB 34 STNKEKL 40

RESULT 13

CH10_MYCBO STANDARD; PRT; 99 AA.
ID CH10_MYCBO
AC P15020;
DT 01-APR-1990 (Rel. 14, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 10 kDa chaperonin (Protein Cpn10) (groES protein) (Immunogenic protein MPB57).
DE MPB57).
GN GROS OR GROES OR MOPB OR MB3452C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN=BCG;
RX MEDLINE=89052868; PubMed=3056744;
RA Yamaguchi R., Matsuo K., Yamazaki A., Nagai S., Terasaka K.,
RA Yamada T.;
RT "Immunogenic protein MPB57 from Mycobacterium bovis BCG: molecular
RT cloning, nucleotide sequence and expression.";
RL FEBS Lett. 240:115-117(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grogdlin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Bartell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- FUNCTION: Binds to Cpn60 in the presence of Mg-ATP and suppresses
CC the ATPase activity of the latter.
CC -!- SUBUNIT: Heptamer of 7 subunits arranged in a ring (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the groES chaperonin family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X13970; CAA32149.1; -;
DR EMBL: M35389; AAA25365.1; -;
DR EMBL: BX248346; CAD95639.1; -;
DR PIR: S01784; BVWY7B.

DR HSSP: P05380; 1AON.
DR HAMAP: MF 00580; -; 1.
DR InterPro: IPR001476; Chaprinin_Cpn10.
DR Pfam: PF00166; cpn10; 1.
DR PRINTS: PR00297; CHAPERONIN10.
DR ProDom: PD00566; Chaprinin_Cpn10; 1.
DR PROSITE: PS00681; CHAPERONINS_CPN10; 1.
KW Chaperone; Antigen; Heat shock; Complete proteome.
FT INIT_MET 0
FT CONFLICT 94 98 LAVVS -> VGRRF (IN REF. 1).
SQ SEQUENCE 99 AA; 10673 MW; 1DD128E75CF19AF7 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 99;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 GEKRIPL 611
DB 54 GEKRIPL 60

RESULT 14

CH10_MYCTU STANDARD; PRT; 99 AA.
ID CH10_MYCTU
AC P09621;
DT 01-MAR-1989 (Rel. 10, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 10 kDa chaperonin (Protein Cpn10) (groES protein) (BCG-A heat shock
DE protein) (10 kDa antigen).
GN GROS OR GROES OR MOPB OR CPN10 OR RV3418C OR MT3527 OR MTCV78.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=89016584; PubMed=2902558;
RA Baird P.N., Hall L.M., Coates A.R.M.;
RT "A major antigen from Mycobacterium tuberculosis which is homologous
RT to the heat shock proteins groES from E. coli and the htpA gene
RT product of *Coxiella burnetii*.";
RL Nucleic Acids Res. 16:9047-9047(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=90095443; PubMed=2480990;
RA Baird P.N., Hall L.M.C., Coates A.R.M.;
RT "Cloning and sequence analysis of the 10 kDa antigen gene of
RT Mycobacterium tuberculosis.";
RL J. Gen. Microbiol. 135:931-939(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Erdmann;
RX MEDLINE=89160258; PubMed=2564178;
RA Shinnick T.M., Plikaytis B.P., Hyche A.D., van Ledingham R.M.,
RA Walker L.L.;
RT "The Mycobacterium tuberculosis BCG-a protein has homology with the
RT *Escherichia coli* GroES protein.";
RL Nucleic Acids Res. 17:1254-1254(1989).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Erdmann;
RX MEDLINE=93219332; PubMed=7681982;
RA Kong T.H., Coates A.R.M., Butcher P.D., Hickman C.J., Shinnick T.M.;
RT "Mycobacterium tuberculosis expresses two chaperonin-60 homologs.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:2608-2612(1993).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce> or send an email to license@isb-sib.ch).

EMBL; AJ309317; CAC38864.2; -;
EMBL; AJ319607; CAC38864.2; JOINED.
EMBL; AJ319608; CAC38864.2; JOINED.
EMBL; AJ319609; CAC38864.2; JOINED.
EMBL; AJ319610; CAC38864.2; JOINED.
EMBL; AJ319611; CAC38864.2; JOINED.
EMBL; AJ319612; CAC38864.2; JOINED.
EMBL; AJ319613; CAC38864.2; JOINED.
EMBL; AK007353; -; NOT_ANNOTATED_CDS.
MGI; MGI:107507; Evpl_1
InterPro; IPR001101; Plectin_repeat.
InterPro; IPR002017; Spectrin.
Pfam; PF00681; Plectin; 3.
Pfam; PF00435; spectrin; 2.
SMART; SM00250; PLEC; 8.
Keratization; Repeat; Coiled coil; Cytoskeleton; Structural protein.
DOMAIN 1 841
DOMAIN 842 1674
DOMAIN 1675 2035
DOMAIN 12 28
REPEAT 229 330
DOMAIN 842 1664
REPEAT 1186 1227
REPEAT 1679 1714
REPEAT 1819 1856
REPEAT 1857 1894
REPEAT 1895 1932
REPEAT 1933 1970
REPEAT 1971 2008
CONFLICT 1861 1864
SEQUENCE 2035 AA; 232317 MW; EB4AID9CAED9641F CRC64;
1.0%; Score 8; DB 1; Length 2035;
Best Local Similarity 100.0%; Pred.No.40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 LKEAEVLL 738 ✓
DB 87 LKEAEVLL 94

RESULT 12
Y872 PYRAB STANDARD; PRT; 70 AA.
AC QSV038;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical UPF0165 protein PYRAB0720.
GN PYRAB0720 OR PAB8218.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -i- SIMILARITY: Belongs to the UPF0165 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Deicher A., Usterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND
 CC PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION (BY SIMILARITY).
 CC -1- SIMILARITY: Belongs to the SMC family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ14609; CAC93884.1; -- INIT.
 CC EMBL; Z74697; CAA98982.1; ALT INIT.
 CC TIGR; MT2990; --
 CC Tuberculin; Rv2922c; --
 CC InterPro; IPR003439; ABC transporter.
 CC InterPro; IPR003405; SMC C.
 CC InterPro; IPR003395; SMC N.
 CC Pfam; PF02483; SMC C; 1.
 CC Pfam; PF02463; SMC N; 1.
 CC ATP-binding; Coiled coil; Complete proteome.
 KW NP BIND 31 38 ATP (POTENTIAL).
 FT DOMAIN 167 289 COILED COIL (POTENTIAL).
 FT DOMAIN 330 499 COILED COIL (POTENTIAL).
 FT DOMAIN 659 842 COILED COIL (POTENTIAL).
 FT DOMAIN 899 929 COILED COIL (POTENTIAL).
 FT DOMAIN 979 1038 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1205 AA; 130637 MW; A3B2A813B58EACF3 CRC64;
 Query Match 1.0%; Score 8; DB 1; Length 1205;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 SELAAAEA 210
 Db 672 SELAAAEA 679
 RESULT 10
 EVPL HUMAN STANDARD; PRT; 2033 AA.
 AC Q2817;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Envoplakin (210 kDa paraneoplastic pemphigus antigen) (p210) (210 kDa
 DE cornified envelope precursor).
 GN EVPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=96326676; PubMed=8707850;
 RA Ruhrberg C., Hajibagheri M.A.N., Simon M., Dooley T.P., Watt F.M.;
 RT "Envoplakin, a novel precursor of the cornified envelope that has
 RT homology to desmoplakin.";
 RL J. Cell Biol. 134:715-729(1996).
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=99339988; PubMed=10409435;
 RA Risk J.M., Ruhrberg C., Hennies H.-C., Mills H.S., Di Colandrea T.,
 RA Evans K.B., Ellis A., Watt F.M., Bishop D.T., Spurr N.K.,
 RA Stevens H.P., Leigh I.M., Reis A., Kelsell D.P., Field J.K.;
 RT "Envoplakin, a possible candidate gene for focal NEPK/esophageal
 RT cancer (TOC): the integration of genetic and physical maps of the TOC
 RT region on 17q25.";
 RL Genomics 59:234-242(1999).
 RN [3]
 CC LIPIDATION.
 RP MEDLINE=98316349; PubMed=9651377;
 RA Marekov L.N., Steinert P.M.;
 RT "Ceramides are bound to structural proteins of the human foreskin
 RT epidermal cornified cell envelope.";
 RL J. Biol. Chem. 273:17763-17770(1998).
 CC -1- FUNCTION: Component of the cornified envelope of keratinocytes.
 CC May link the cornified envelope to desmosomes and intermediate
 CC filaments.
 CC -1- SUBUNIT: May form a homodimer or a heterodimer with PPL.
 CC -1- SUBCELLULAR LOCATION: COLOCALIZED WITH DSP AT DESMOSOMES AND ALONG
 CC INTERMEDIATE FILAMENTS.
 CC -1- TISSUE SPECIFICITY: Exclusively expressed in stratified squamous
 CC epithelia.
 CC -1- INDUCTION: During differentiation of epidermal keratinocytes.
 CC -1- PTM: Substrate of transglutaminase. Some glutamines and lysines
 CC are cross-linked to other envoplakin molecules, to other proteins
 CC such as keratin, desmoplakin, periplakin and involucrin, and to
 CC lipids like omega-hydroxyceramide.
 CC -1- SIMILARITY: Contains 7 plectrin repeats.
 CC -1- SIMILARITY: Contains 1 spectrin repeat.
 CC -1- SIMILARITY: Belongs to the plakin or cytolinker family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U53786; AAC64662.1; --
 CC EMBL; U72849; AAD00186.1; --
 CC EMBL; U72843; AAD00186.1; JOINED.
 CC EMBL; U72845; AAD00186.1; JOINED.
 CC EMBL; U72846; AAD00186.1; JOINED.
 CC EMBL; U72847; AAD00186.1; JOINED.
 CC EMBL; U72848; AAD00186.1; JOINED.
 CC Genew; HGNC:3503; EVPL.
 CC NIM; 601590; --
 CC GO; GO:0009506; C:plasmodesma; TAS.
 CC GO; GO:0005198; F:structural molecule activity; TAS.
 CC GO; GO:0008544; P:epidermal differentiation; TAS.
 CC InterPro; IPR001101; Plectrin_repeat.
 CC Pfam; PF00681; Plectrin; 3.
 CC SMART; SM00250; PLEC; 8.
 CC Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein;
 KW Lipoprotein.
 KW DOMAIN 1 841 GLOBULAR 1.
 FT DOMAIN 842 1673 CENTRAL FIBROUS ROD DOMAIN.
 FT DOMAIN 1674 2033 GLOBULAR 2.
 FT DOMAIN 12 28 4 X 4 AA TANDEM REPEATS OF K-G-S-P.
 FT REPEAT 229 330 SPECTRIN.
 FT DOMAIN 845 1135 COILED COIL (POTENTIAL).
 FT REPEAT 1185 1226 PLECTIN 1.
 FT REPEAT 1678 1713 PLECTIN 2.
 FT REPEAT 1818 1855 PLECTIN 3.
 FT REPEAT 1856 1893 PLECTIN 4.
 FT REPEAT 1894 1931 PLECTIN 5.
 FT REPEAT 1932 1969 PLECTIN 6.
 FT REPEAT 1970 2007 PLECTIN 7.
 FT LIPID 657 657 Omega-hydroxyceramide glutamate ester
 FT (Potential).
 FT

RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahler C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.H., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kammel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
RN [3]
RN PARTIAL SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC TISSUE=Embryo;
RX MEDLINE=98198453; PubMed=9531534;
RA Steneberg P., England C., Kronham J., Weaver T.A., Samakovlis C.;
RT "Translational readthrough in the hdc mRNA generates a novel branching
inhibitor in the *Drosophila* trachea";
RL Genes Dev. 12:956-967(1998).
CC -!- FUNCTION: Required for imaginal cell differentiation, may be
involved in hormonal responsiveness during metamorphosis. Involved
in an inhibitory signaling mechanism to determine the number of
cells that will form unicellular sprouts in the trachea. Regulated
by transcription factor seg. The longer hdc protein is completely
functional and the shorter protein carries some function.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Expressed in all imaginal cells of the embryo
and larvae. Expressed in a subset of tracheal fusion cells from
stage 14 to the end of embryogenesis in metameres 2-9, lateral
trunk and ventral anastomoses.
CC -!- MISCELLANEOUS: Readthrough of the terminator UAA occurs between
codons for Ala-650 and His-652. Readthrough is not always
suppressed as the shorter protein is more abundant.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; Z50097; CAA90425.1; -;
CC EMBL; Z50097; CAB58233.1; -;
CC EMBL; AE003773; AAF57033.1; ALT_SEQ.
CC FlyBase; FBgn0010113; hdc.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0007430; P:terminal branching of trachea, cytoplasmic . . .; NAS.
CC KW Developmental protein.
CC CHAIN 1 1080 HEADCASE PROTEIN.

PT CHAIN 1 650 HEADCASE SHORT PROTEIN.
FT DOMAIN 57 66 POLY-GLY.
FT DOMAIN 211 218 POLY-ASN.
FT DOMAIN 219 227 POLY-GLY.
FT DOMAIN 343 350 POLY-GLN.
FT DOMAIN 381 395 POLY-GLN.
FT DOMAIN 723 769 GLN-RICH.
FT DOMAIN 801 815 POLY-GLN.
FT DOMAIN 845 854 POLY-SER.
FT DOMAIN 887 891 POLY-SER.
FT DOMAIN 965 970 POLY-SER.
FT DOMAIN 1030 1036 POLY-SER.
FT CONFLICT 85 85 H -> P (IN REF. 1).
FT CONFLICT 190 191 PT -> SN (IN REF. 1).
FT CONFLICT 226 226 A -> G (IN REF. 1).
FT CONFLICT 243 244 SY -> HD (IN REF. 1).
FT CONFLICT 279 310 SGVLTSLATSLTIRMSNLTNNVGLDLRAGS -> PACCR
PVRPLSATSSIRTSMPGPARGQWQ (IN REF. 1).
FT CONFLICT 342 342 P -> A (IN REF. 1).
FT CONFLICT 353 353 L -> V (IN REF. 1).
FT CONFLICT 383 383 Q -> P (IN REF. 1).
FT CONFLICT 432 432 D -> E (IN REF. 1).
FT CONFLICT 641 641 T -> S (IN REF. 1).
FT CONFLICT 695 695 P -> Q (IN REF. 1).
FT CONFLICT 852 852 S -> SS (IN REF. 1).
FT CONFLICT 1067 1067 A -> R (IN REF. 1).
SQ SEQUENCE 1080 AA; 117446 MW; 87EB144BA0D1B787 CRC64;
Query Match 1.0%; Score 8; DB 1; Length 1080;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 247 TNYTNTSN 254 ✓
DB 836 TNYTNTSN 843
RESULT 9
SMC_MYCTU STANDARD; PRT; 1205 AA.
ID - SMC_MYCTU
AC Q10970;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chromosome partition protein SMC.
GN SMC OR RV2922C OR M2950 OR MIC338.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Cobbe N., Heck M.M.S.;
RT "Phylogenetic analysis of SMC proteins.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.

entities requires a license agreement (See <http://www.isb-sib.ch/and/or send an email to license@isb-sib.ch>).

EMBL; AEO00050; AAB96152.1; -
PIR; S73830; S73830.
MEROPS; S16.004; -
InterPro; IPR003593; AAA_ATPase.
InterPro; IPR003959; AAA_ATPase_central.
InterPro; IPR008269; Pept_S16_C.
InterPro; IPR004815; Pept_S16_Lon.
InterPro; IPR003111; Pept_S16_N.
InterPro; IPR008268; Peptid_S16_AS.
InterPro; IPR001984; Peptidase_S16.
Pfam; PF00004; AAA; 1.
Pfam; PF02190; Lon; 1.
Pfam; PF05362; Lon_C; 1.
PRINTS; PR00830; ENDOLAPTASE.
SMART; SMO0382; AAA; 1.
SMART; SMO0464; Lon; 1.
TIGRFAMs; TIGR00763; Lon; 1.
PROSITE; PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
FT DOMAIN 304 308 POLY-SER
FT NP BIND 379 386 ATP (POTENTIAL).
FT ACT_SITE 702 702 BY SIMILARITY.
SQ SEQUENCE 795 AA; 9203 MW; 7E7855082060C891 CRC64;
Query Match 1.0%; Score 8; DB 1; Length 795;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 466 ALDKLLER 473 ✓
|||||||
Db 127 ALDKLLER 134
RESULT 7
Y240 ARATH
ID Y240 ARATH STANDARD; PRT; 861 AA.
AC Q22224; Q94CF2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Protein At2g41620.
GN At2g41620 OR T32G6.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euraids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gail J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Nierman W.C., White O., Bisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,

```

DR GO: 0005634; C:nucleus; TAS.
DR GO: 0003704; F:specific RNA polymerase II transcription fa. . . ; TAS.
DR GO: 0000122; P:negative regulation of transcription from P. . . ; TAS.
DR GO: 0000615; P:response to viruses; TAS.
DR InterPro: IPR001346; IRF.
DR InterPro: IPR008984; SMAD_FHA.
DR Pfam: PF00605; IRF; 1.
DR PRINTS: PR00267; INTFRNREGFCT.
DR ProDom: PD002355; IRF; 1.
DR SMART: SM00348; IRF; 1.
DR PROSITE: PS00601; IRF; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Activator;
KW Alternative splicing.
FT DNA BIND 13 122 TRYPTOPHAN PENTAD REPEAT
FT VARSPPLIC 1 6 MALAPE -> MPVPERPAAGDSPPGTR (in isoform
FT D).
FT VARSPPLIC 152 164 /FTID=VSP_002757.
FT VARSPPLIC 165 503 /GGPPGFLAHTHA -> AQGSLGSGCTGGQ (in
FT VARSPPLIC 228 256 /FTID=VSP_002758.
FT VARSPPLIC 179 179 Missing (in isoform C).
FT CONFLICT 179 179 Missing (in isoform B).
FT CONFLICT 412 412 /FTID=VSP_002759.
FT CONFLICT 412 412 /FTID=VSP_002760.
FT CONFLICT 412 412 E -> K (IN REF. 2).
FT CONFLICT 412 412 Q -> R (IN REF. 3).
SQ SEQUENCE 503 AA; 54278 MW; AA6A39E0E272727C CRC64;

Query Match 1.0%; Score 8; DB 1; Length 503;
Best Local Similarity 100.0%; Pred.No.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 357 PSPGPQPA 364
Db 244 PSPGPQPA 251
|||||

RESULT 6
LON_MYCPN STANDARD; PRT; 795 AA.
AC P78025;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease Ia (EC 3.4.21.53).
GN LON OR MFN332 OR MF504.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Hamelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herimann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- FUNCTION: DEGRADES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS
CC IN PRESENCE OF ATP. DEGRADES THE REGULATORY PROTEINS RCSA AND
CC SULA. HYDROLYZES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED IN THE
CC PROTEIN SUBSTRATE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to peptidase family S16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC GO: 0005737; C:cytoplasm; TAS.

```

NR SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE=94303173; PubMed=8030224;
RT Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus";
RL Virology 202:586-605(1994).
CC -!- SIMILARITY: TO CORRESPONDING ORF IN OPMNPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L22858; AAA66659.1; -.
DR PIR; E72853; E72853.
KW Hypothetical protein.
SQ SEQUENCE 71 AA; 8569 MW; ADFF85AC68E16DD3 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 EEINRQKQ 137 ✓
DB 20 EEINRQKQ 27

RESULT 3
SC17_NEUCR STANDARD; PRT; 292 AA.
AC Q9P6A5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable vesicular-fusion protein sc17 homolog.
GN B1D.150.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A / FGSC 987;
RX MEDLINE=22542210; PubMed=12655011;
RA Manthaupt G., Montrone C., Haase D., Mewes H.-W., Aign V.,
RA Hoheisel J.D., Fattmann B., Nyakatura G., Kempken F., Maier J.,
RA Schulte U.;
RT "What's in the genome of a filamentous fungus? Analysis of the
RT Neurospora genome sequence."
RL Nucleic Acids Res. 31:1944-1954(2003).
CC -!- FUNCTION: Required for vesicular transport between the endoplasmic
CC reticulum and the Golgi apparatus (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic peripheral membrane protein (By
CC similarity).
CC -!- SIMILARITY: Belongs to the SNAP family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL355927; CAB91264.1; -.
DR PIR; T49361; T49361.
DR InterPro; IPR000744; NSF_attach.
DR InterPro; IPR008941; TPR-like.

DR Pfam; PF02071; NSF; 2.
DR PRINTS; PR00448; NSFATTACHMNT.
KW Transport; Protein transport; Endoplasmic reticulum; Golgi stack.
SQ SEQUENCE 292 AA; 32805 MW; C9BD75DB128E19E7 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 292;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 NDGAVALA 157 ✓
DB 150 NDGAVALA 157

RESULT 4
AAC2_DICDI STANDARD; PRT; 448 AA.
ID AAC2_DICDI
AC F14136;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE AAC-rich mRNA clone AAC11 protein (Fragment).
DE AAC11.
GN Dictyostelium discoideum (Slime mold).
OS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=90066348; PubMed=2511421;
RA Shaw D.R., Richter H., Giorda R., Ohmachi T., Ennis H.L.;
RT "Nucleotide sequences of Dictyostelium discoideum developmentally
RT regulated cDNAs rich in (AAC) imply proteins that contain clusters of
RT asparagine, glutamine, or threonine."
RL Mol. Gen. Genet. 218:453-459(1989).
CC -!- DEVELOPMENTAL STAGE: The concentration of AAC-rich mRNAs is low
CC in dormant spores and growing cells, but increases during
CC spore-germination and multicellular development.
CC -!- MISCELLANEOUS: Several proteins derive from AAC-rich mRNA, which,
CC due to a frameshift also have ACA and CAA codons and thus are
CC Asn-, Thr- or Gln-rich.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16522; CAA34529.1; -.
DR PIR; S05355; S05355.
DR DictyBase; DDB0001892; AAC11.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR000116; Highmoblty_IY.
DR Pfam; PF02178; AT_hook; 4.
DR PRINTS; PR00929; ATHOOK.
DR ProDom; PD005593; Highmoblty_IY; 1.
DR SMART; SM00384; AT_hook; 4.
KW DNA-binding; Repeat.
FT NON TER 1
FT DOMAIN 5 25 GLN-RICH.
FT DOMAIN 115 144 ASN-RICH.
FT DOMAIN 268 384 ASN-RICH.
SQ SEQUENCE 448 AA; 48636 MW; D8FD80D910D99817 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 448;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 NTSNNSNT 257 ✓
DB 378 NTSNNSNT 385

CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CC NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF 519-540; 597-616
 RP AND 857-874.
 RC TISSUE=Embryo;
 RX MEDLINE=93145326; PubMed=7678780;
 RA Hoey T., Weinzierl R.O.J., Gill G., Chen J.-L., Dynlacht B.D.,
 RA Tjian R.;
 RT "Molecular cloning and functional analysis of Drosophila TAF110
 RT reveal properties expected of coactivators.";
 RL Cell 72:247-260(1993).
 [2]
 RP SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF 398-406; 520-540
 RP AND 860-877.
 RC TISSUE=Embryo;
 RX MEDLINE=93317591; PubMed=8327460;
 RA Kokubo T., Gong D.-W., Roeder R.G., Horikoshi M., Nakatani Y.;
 RT "The Drosophila 110-kDa transcription factor TFIID subunit directly
 RT interacts with the N-terminal region of the 230-kDa subunit.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.-H., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M.C., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [4]
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).
 RC STRAIN=Berkley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarini H., Kronmiller B., Pacle J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A Drosophila full-length cDNA resource.";

Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 -!- FUNCTION: TFIID is a multimeric protein complex that plays a
 central role in mediating promoter responses to various activators
 and repressors. May function as a coactivator by serving as a site
 of protein-protein contact between activators like Sp1 (or btd)
 and TFIID complex.
 -!- SUBUNIT: TFIID is composed of TATA binding protein (TBP) and a
 number of TBP-associated factors (TAFs).
 -!- SUBCELLULAR LOCATION: Nuclear.
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Name=Long;
 IsoId=P47825-1; Sequence=Displayed;
 Name=Short;
 IsoId=P47825-2; Sequence=VSP_004441;
 Note=No experimental confirmation available;
 -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 -!- SIMILARITY: Contains 1 TAFH/NHR1 domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; L06861; -; NOT ANNOTATED_CDS.
 EMBL; S63550; AAB27433.1; -.
 EMBL; AE003528; AAF49536.1; -.
 EMBL; AY069807; AAL39952.1; -.
 PIR; A48184; A48184.
 TRANSEAC; T02121; -.
 FlyBase; FBgn0010280; Taf4.
 GO; GO:0005669; C:transcription factor TFIID complex; IPI.
 GO; GO:0016251; F:General RNA polymerase II transcription fac...; IPI.
 GO; GO:0006355; P:regulation of transcription, DNA-dependent; IPI.
 GO; GO:0006367; P:transcription initiation from Pol II promoter; IPI.
 InterPro; IPR007900; TAF4.
 InterPro; IPR003894; TAF_Hom.
 Pfam; PF05236; TAF4; 1.
 SMART; SM00549; TAFH; 1.
 Transcription regulation; Nuclear protein; Alternative splicing.
 DOMAIN 293 383 TAFH/NHR1.
 DOMAIN 66 82 POLY-GLN.
 DOMAIN 108 111 POLY-GLN.
 DOMAIN 259 265 POLY-GLN.
 VARSPLIC 138 207 Missing (in isoform Short).
 CONFLICT 114 114 Missing (VSP_004441).
 SEQUENCE 921 AA; 99338 MW; 27E6852859872767 CRC64;
 Query Match 1.1%; Score 9; DB 1; Length 921;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 249 TTTSSNSNT 257 ✓
 Db 178 TTTSSNSNT 186
 RESULT 2
 Y029 NPVAC STANDARD; PRT; 71 AA.
 ID Y029 NPVAC
 AC P41433;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE Hypothetical 8.6 kDa protein in IAP1-SOD intergenic region.
 OS Autographa californica nuclear polyhedrosis virus (ACNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OX NCBI_TaxID=46015;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:17:54 ; Search time 19 Seconds
(without alignments)
2181.465 Million cell updates/sec

Title: US-09-765-271-56
Perfect score: 796
Sequence: 1 SYELGLYQARTVKENRVSY.....KLALIKGNSPSSVSKEKIN 796

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 9 | 1.1 | 921 | 1 | T2D3 DROME |
| 2 | 8 | 1.0 | 71 | 1 | Y029 NPVAC |
| 3 | 8 | 1.0 | 292 | 1 | SC17 NEUCR |
| 4 | 8 | 1.0 | 448 | 1 | AAC2 DICDI |
| 5 | 8 | 1.0 | 503 | 1 | IR77 HUMAN |
| 6 | 8 | 1.0 | 795 | 1 | ION MYCPN |
| 7 | 8 | 1.0 | 861 | 1 | Y240 ARATH |
| 8 | 8 | 1.0 | 1080 | 1 | HDC DROME |
| 9 | 8 | 1.0 | 1205 | 1 | SMC MYCTU |
| 10 | 8 | 1.0 | 2033 | 1 | EVPL HUMAN |
| 11 | 8 | 1.0 | 2035 | 1 | EVPL MOUSE |
| 12 | 7 | 0.9 | 70 | 1 | Y872 PYRAB |
| 13 | 7 | 0.9 | 99 | 1 | CH10 MYCBO |
| 14 | 7 | 0.9 | 99 | 1 | CH10 MYCTU |
| 15 | 7 | 0.9 | 101 | 1 | GLR1 SCHPO |
| 16 | 7 | 0.9 | 101 | 1 | RS16 UREPA |
| 17 | 7 | 0.9 | 106 | 1 | YQCC HAEIN |
| 18 | 7 | 0.9 | 115 | 1 | RNPA STAEP |
| 19 | 7 | 0.9 | 121 | 1 | INSC ECOLI |
| 20 | 7 | 0.9 | 121 | 1 | INSC SHIFL |
| 21 | 7 | 0.9 | 124 | 1 | RNP ANTAM |
| 22 | 7 | 0.9 | 143 | 1 | Y082 RICPR |
| 23 | 7 | 0.9 | 148 | 1 | PTVA ECOLI |
| 24 | 7 | 0.9 | 154 | 1 | MUTT STRAM |
| 25 | 7 | 0.9 | 158 | 1 | GREX STRAM |
| 26 | 7 | 0.9 | 158 | 1 | Y052 AKCFU |
| 27 | 7 | 0.9 | 166 | 1 | GC51 ARATH |
| 28 | 7 | 0.9 | 174 | 1 | HUNE DROMU |
| 29 | 7 | 0.9 | 177 | 1 | APT CHLTE |
| 30 | 7 | 0.9 | 178 | 1 | NUSG NEIMA |
| 31 | 7 | 0.9 | 179 | 1 | APT HELPU |
| 32 | 7 | 0.9 | 185 | 1 | VATE DEIRA |
| 33 | 7 | 0.9 | 197 | 1 | CYCL METEX |

RESULT 1

T2D3 DROME STANDARD; Q8T9E0; Q9VUY7; PRT; 921 AA.
ID T2D3 DROME AC P47825; P49845; Q8T9E0; Q9VUY7;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription initiation factor TFIID 110 kDa subunit (P110)
DE (TAFII-110) (110 kDa TBP-associated factor).
GN TAF4 OR TAF110 OR CG5444.
OS Drosophila melanogaster (Fruit fly).

ALIGNMENTS

| | | | | | | |
|----|---|-----|-----|---|------------|--------------------|
| 34 | 7 | 0.9 | 201 | 1 | RB1B HUMAN | O9hou4 homo sapien |
| 35 | 7 | 0.9 | 201 | 1 | RB1B RAT | P10536 rattus norv |
| 36 | 7 | 0.9 | 210 | 1 | APSI_SCHPO | Q09790 schizosacch |
| 37 | 7 | 0.9 | 211 | 1 | L4P_ARATH | O64628 arabidopsis |
| 38 | 7 | 0.9 | 211 | 1 | YRAR ECOLI | P45469 escherichia |
| 39 | 7 | 0.9 | 213 | 1 | CPCF SYNEL | P50038 synechococc |
| 40 | 7 | 0.9 | 213 | 1 | URE1 HELMU | P50044 helicobacte |
| 41 | 7 | 0.9 | 238 | 1 | PSE2 RAT | O63798 rattus norv |
| 42 | 7 | 0.9 | 239 | 1 | PSE2 MOUSE | P97372 mus musculu |
| 43 | 7 | 0.9 | 271 | 1 | EL2 RAT | P00774 rattus norv |
| 44 | 7 | 0.9 | 280 | 1 | FPG OCEIH | Q8epe6 oceanobacil |
| 45 | 7 | 0.9 | 284 | 1 | LICH VIBCH | O07350 vibrio chol |
| 46 | 7 | 0.9 | 284 | 1 | TPM1 BIOGL | P42636 biophalari |
| 47 | 7 | 0.9 | 284 | 1 | TPM2 BIOGL | P43689 biophalari |
| 48 | 7 | 0.9 | 284 | 1 | TPM1 TRICO | P15846 trichostrom |
| 49 | 7 | 0.9 | 293 | 1 | LECI CLALU | Q39528 cladastis |
| 50 | 7 | 0.9 | 311 | 1 | OYJ3 HUMAN | O8nh60 homo sapien |
| 51 | 7 | 0.9 | 316 | 1 | TAL EUCAP | O8ka27 buchnera ap |
| 52 | 7 | 0.9 | 318 | 1 | XERC LEPIN | Q7zam8 leptospira |
| 53 | 7 | 0.9 | 323 | 1 | PTPA RABIT | Q28717 oryctolagus |
| 54 | 7 | 0.9 | 327 | 1 | FIP1 YEAST | P45976 saccharomyc |
| 55 | 7 | 0.9 | 327 | 1 | KAPR DICDI | P05987 dictyostell |
| 56 | 7 | 0.9 | 338 | 1 | G3P2 AGABI | P32636 agaricus bi |
| 57 | 7 | 0.9 | 338 | 1 | G3P SCHMA | P20287 schistosoma |
| 58 | 7 | 0.9 | 344 | 1 | DCUP AGRT5 | Q8ub16 agrobacteri |
| 59 | 7 | 0.9 | 358 | 1 | PTPA HUMAN | Q15257 homo sapien |
| 60 | 7 | 0.9 | 368 | 1 | LEMT CLABR | O04385 clarkia bre |
| 61 | 7 | 0.9 | 386 | 1 | PHEA ECOLI | P07022 escherichia |
| 62 | 7 | 0.9 | 387 | 1 | YHFY ECOLI | P45550 escherichia |
| 63 | 7 | 0.9 | 394 | 1 | PGK THETN | O8r965 thermocanae |
| 64 | 7 | 0.9 | 397 | 1 | PGK CHLTE | O8kae1 chlorobium |
| 65 | 7 | 0.9 | 405 | 1 | IF5 YEAST | P18431 saccharomyc |
| 66 | 7 | 0.9 | 415 | 1 | RPA3 YEAST | Q01080 saccharomyc |
| 67 | 7 | 0.9 | 416 | 1 | HEM1 AERPE | Q9y972 aeropyrum p |
| 68 | 7 | 0.9 | 424 | 1 | YI2A ECOLI | P51026 escherichia |
| 69 | 7 | 0.9 | 430 | 1 | TRPB HALN1 | Q9hsc0 halobacteri |
| 70 | 7 | 0.9 | 432 | 1 | GSA AERPE | Q9y919 aeropyrum p |
| 71 | 7 | 0.9 | 446 | 1 | TBG SCHJP | Q9y982 schizosacch |
| 72 | 7 | 0.9 | 446 | 1 | TBG SCHPO | P25295 schizosacch |
| 73 | 7 | 0.9 | 447 | 1 | GCSA BACHD | Q9X935 bacillus ha |
| 74 | 7 | 0.9 | 450 | 1 | V50K BYDVP | P09516 barley yell |
| 75 | 7 | 0.9 | 451 | 1 | TBGI HUMAN | P23258 homo sapien |
| 76 | 7 | 0.9 | 451 | 1 | TBGI MOUSE | Q9z310 mus musculu |
| 77 | 7 | 0.9 | 451 | 1 | TBGI HUMAN | Q9nrb3 homo sapien |
| 78 | 7 | 0.9 | 451 | 1 | TBGI MOUSE | Q8vck3 mus musculu |
| 79 | 7 | 0.9 | 451 | 1 | TBG XENLA | P23330 xenopus lae |
| 80 | 7 | 0.9 | 455 | 1 | YXX5 CAEL | Q18179 caenorhabdi |
| 81 | 7 | 0.9 | 460 | 1 | SELA PASMU | Q9ck66 pasteurella |
| 82 | 7 | 0.9 | 461 | 1 | TBGI EUPOC | P34786 euplotes oc |
| 83 | 7 | 0.9 | 461 | 1 | TBGI EUPOC | P90548 euplotes oc |
| 84 | 7 | 0.9 | 461 | 1 | TBG NEUCR | P53377 neurospora |
| 85 | 7 | 0.9 | 461 | 1 | TRIB HUMAN | P20333 homo sapien |
| 86 | 7 | 0.9 | 462 | 1 | TBGI EUPCR | P54403 euplotes cr |
| 87 | 7 | 0.9 | 467 | 1 | V51K BWYF | P09514 beet wester |
| 88 | 7 | 0.9 | 471 | 1 | V51K BWYVG | P09515 beet wester |
| 89 | 7 | 0.9 | 476 | 1 | YHGF NEIGO | Q51062 neisseria g |
| 90 | 7 | 0.9 | 479 | 1 | VGLC HSV2G | P03173 herpes simp |

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 SELAAEA 210

Db 756 SELAAEA 763

Search completed: October 1, 2004, 07:33:26
Job time : 33 secs

C;Genetics:
 A;Gene: lon; F10 orf795
 A;Genetic code: SGC3
 C;Superfamily: ATP-dependent Lon protease
 C;Keywords: ATP; DNA binding; heat shock; hydrolase; nucleotide binding; P-loop; serine
 F;379-386/Region: nucleotide-binding motif A (P-loop)
 F;442-447/Region: nucleotide-binding motif B
 F;702/Active site: Ser #status predicted

Query Match 1.0%; Score 8; DB 1; Length 795;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 ALDKLLER 473
 |||||
 DB 127 ALDKLLER 134

RESULT 22
 T00818
 hypothetical protein At2g41620 [imported] - Arabidopsis thaliana
 N;Alternate names: hypothetical protein T32G6.14
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
 C;Accession: T00818; A84844
 R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
 submitted to the EMBL Data Library, November 1997
 A;Description: Arabidopsis thaliana chromosome II BAC T32G6 genomic sequence.
 A;Reference number: Z14163
 A;Accession: T00818
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-825 <R0U>
 A;Cross-references: EMBL:AC002510; NID:g2618683; PID:g2618698
 A;Experimental source: cultivar Columbia
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: A84844
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-825 <STO>
 A;Cross-references: GB:AE002093; NID:g2618698; PIDN:AAB84345.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: At2g41620; T32G6.14
 A;Map position: 2
 A;Introns: 43/3; 74/2; 96/3; 116/3; 146/3; 234/3; 266/3; 311/3; 329/3; 368/3; 669/3; 729
 A;Note: T32G6.14
 C;Superfamily: Arabidopsis thaliana hypothetical protein At2g41620

Query Match 1.0%; Score 8; DB 2; Length 825;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 REGINAEQ 48
 |||||
 DB 78 REGINAEQ 85

RESULT 23
 C75313
 probable proteinase - Deinococcus radiodurans (strain R1)
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C;Accession: C75313
 R;White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: C75313
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-840 <WHI>
 A;Cross-references: GB:AE002047; GB:AE000513; NID:g6459915; PIDN:AAF11680.1; PID:g645992
 F;442-447/Region: nucleotide-binding motif A (P-loop)
 F;702/Active site: Ser #status predicted

Query Match 1.0%; Score 8; DB 2; Length 840;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 DGLVFDPA 294
 |||||
 DB 332 DGLVFDPA 339

RESULT 24
 A64714
 helicase - Helicobacter pylori (strain 26695)
 C;Species: Helicobacter pylori
 C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C;Accession: A64714
 R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney,
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
 Nature 388, 539-547, 1997
 A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A;Reference number: A64520; MUID:97394467; PMID:9252185
 A;Accession: A64714
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-945 <TOM>
 A;Cross-references: GB:AE000653; GB:AE000511; NID:g2314733; PIDN:AAD08593.1; PID:g2314737

Query Match 1.0%; Score 8; DB 2; Length 945;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 ERLGKENS 506
 |||||
 DB 657 ERLGKENS 664

RESULT 25
 B70748
 probable smc protein - Mycobacterium tuberculosis (strain H37RV)
 C;Species: Mycobacterium tuberculosis
 C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C;Accession: B70748
 R;Cole, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A;Reference number: A70500; MUID:98295987; PMID:9634230
 A;Accession: B70748
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-1289 <COL>
 A;Cross-references: GB:Z74697; GB:AL123456; NID:g3261602; PIDN:CAA98982.1; PID:gl405958
 A;Experimental source: strain H37RV
 C;Genetics:
 A;Gene: smc
 C;Superfamily: chromosome segregation protein SMCI

Query Match 1.0%; Score 8; DB 2; Length 1289;
 Best Local Similarity 100.0%; Pred. No. 60;

Query Match 1.0%; Score 8; DB 2; Length 292;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 NDGAVALA 157
|||||
Db 150 NDGAVALA 157

RESULT 17
T27791
hypothetical protein ZK228.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T27791
R;Basham, V.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z20419
A;Accession: T27791
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-294 <MIL>
A;Cross-references: EMBL:Z82086; PIDN:CAB04996.1; GSPDB:GN00023; CESP:ZK228.4
A;Experimental source: clone ZK228
C;Genetics:
A;Gene: CESP:ZK228.4
A;Map position: 5
A;Introns: 38/1; 99/1; 143/3; 181/3; 250/2
C;Superfamily: Caenorhabditis elegans hypothetical protein F31F7.1

Query Match 1.0%; Score 8; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRV 18
|||||
Db 262 TVKENNRV 269

RESULT 18
S05355
hypothetical protein (clone AAC11) - slime mold (Dictyostelium discoideum) (fragment)
C;Species: Dictyostelium discoideum
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 29-Oct-1999
C;Accession: S05355
R;Shaw, D.R.; Richter, H.; Giorda, R.; Ohmachi, T.; Ennis, H.L.
Mol. Gen. Genet. 218, 453-459, 1989
A;Title: Nucleotide sequences of Dictyostelium discoideum developmentally regulated cDNA
A;Reference number: S05355; MUID:90066348; PMID:2511421
A;Accession: S05355
A;Molecule type: mRNA
A;Residues: 1-448 <SHA>
A;Cross-references: EMBL:X16522; NID:g7173; PIDN:CAA34529.1; PID:g7174

Query Match 1.0%; Score 8; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 NTSNNSNT 257
|||||
Db 378 NTSNNSNT 385

RESULT 19
F82203
probable multidrug resistance protein VCI409 [imported] - Vibrio cholerae (strain Ni6961)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: F82203
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermlaeva, M.D.; Vanathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: F82203
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-484 <HEI>
A;Cross-references: GB:AE004219; GB:AE003852; NID:g9555899; PIDN:AAF94566.1; GSPDB:GN0012
A;Experimental source: serogroup O1; strain Ni6961; biotype El Tor
C;Genetics:
A;Gene: VCI409
A;Map position: 1

Query Match 1.0%; Score 8; DB 2; Length 484;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 SELAAAEA 210
|||||
Db 169 SELAAAEA 176

RESULT 20
B89957
hypothetical protein ptAA [imported] - *Staphylococcus aureus* (strain N315)
C;Species: *Staphylococcus aureus*
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: B89957
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi,
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; H
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A;Reference number: B89758; MUID:21311952; PMID:11418146
A;Accession: B89957
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-488 <KUR>
A;Cross-references: GB:BA000018; PID:g13701521; PIDN:BAB42815.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: ptAA
C;Superfamily: phosphotransferase system glucose-specific enzyme II, factor II; phosphot

Query Match 1.0%; Score 8; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 NDGAVALA 157
|||||
Db 72 NDGAVALA 79

RESULT 21
S73830
endopeptidase La (EC 3.4.21.53) - *Mycoplasma pneumoniae* (strain ATCC 29342)
N;Alternate names: ATP-dependent proteinase lon; ATP-dependent serine proteinase La; hylp
N;Contains: adenosinetriphosphatase (EC 3.6.1.3)
C;Species: *Mycoplasma pneumoniae*
A;Variety: ATCC 29342
C;Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 02-Jun-2003
C;Accession: S73830
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*.
A;Reference number: S73327; MUID:97105885; PMID:8948633
A;Accession: S73830
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-795 <HIM>
A;Cross-references: EMBL:AB000050; GB:U00089; NID:g1674197; PIDN:AAB96152.1; PID:g1674196
A;Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Comment: This enzyme catalyzes the hydrolysis of large proteins in the presence of ATP

Cell 72, 247-260, 1993
 A:Title: Molecular cloning and functional analysis of Drosophila TAF110 reveal properties
 A:Reference number: A45183; MUID:93145326; PMID:7678780
 A:Accession: A45183
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Note: sequence extracted from NCBI backbone (NCBIP:123832)
 A:Experimental source: embryo
 A:Genetics:
 A:Gene: FlyBase:Taf110
 A:Cross-references: FlyBase:FBgn0010280

Query Match 1.1%; Score 9; DB 2; Length 921;
 Best Local Similarity 100.0%; Pred. No. 4.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TNTSNNST 257
 DB 178 TNTSNNST 186
 |||||

RESULT 13
 E72853
 A:Title: Autographa californica nuclear polyhedrosis virus
 A:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV
 A:Note: dsDNA virus
 A:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 12-Nov-1999
 A:Accession: E72853
 R:AYres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.
 Virology 202, 586-605, 1994
 A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.
 A:Reference number: A72850; MUID:94303173; PMID:8030224
 A:Accession: E72853
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-71 <AY>
 A:Cross-references: GB:L22858; NID:g510708; PIDN:AAA66659.1; PID:g559098
 A:Genetics:
 A:Gene: AcOrf-29

Query Match 1.0%; Score 8; DB 2; Length 71;
 Best Local Similarity 100.0%; Pred. No. 4.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 BEINRQKQ 137
 DB 20 BEINRQKQ 27
 |||||

RESULT 14
 S58921
 tropomyosin isoform TMI - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 20-Jun-2000
 A:Accession: S58921; S62127; S58918
 R:Kagawa, H.; Sugimoto, K.; Matsumoto, H.; Inoue, T.; Imadzu, K.; Sakube, Y.
 J. Mol. Biol. 251, 603-613, 1995
 A:Title: Genome structure, mapping and expression of the tropomyosin gene tmy-1 of Caenorhabditis elegans
 A:Reference number: S58918; MUID:95395840; PMID:7666414
 A:Accession: S58921
 A:Molecule type: mRNA
 A:Residues: 1-284 <KAG>
 A:Cross-references: EMBL:D38540; NID:g1208412; PIDN:BAA07543.1; PID:g1208413
 A:Accession: S62127
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-276, 'NLP', 280-284 <KAF>
 A:Cross-references: EMBL:D38539; NID:g871835; PIDN:BAA07540.1; PID:g1208409
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994
 A:Accession: S58918
 A:Molecule type: protein
 A:Residues: 11-49; 171-211 <KAW>

C:Genetics:
 A:Gene: tmy-1
 A:Introns: 15/3; 61/3; 80/3; 125/2; 164/3; 188/2; 234/3; 258/1
 C:Superfamily: tropomyosin
 C:Keywords: alternative splicing

Query Match 1.0%; Score 8; DB 2; Length 284;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 727 VEAQLKEA 734
 DB 144 VEAQLKEA 151
 |||||

RESULT 15
 S58922
 tropomyosin isoform TMI - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
 A:Accession: S58922; S62128; S58918
 R:Kagawa, H.; Sugimoto, K.; Matsumoto, H.; Inoue, T.; Imadzu, H.; Takuwa, K.; Sakube, Y.
 J. Mol. Biol. 251, 603-613, 1995
 A:Title: Genome structure, mapping and expression of the tropomyosin gene tmy-1 of Caenorhabditis elegans
 A:Reference number: S58918; MUID:95395840; PMID:7666414
 A:Accession: S58922
 A:Molecule type: mRNA
 A:Residues: 1-284 <KAG>
 A:Cross-references: EMBL:D38541; NID:g871837; PIDN:BAA07544.1; PID:g1208414
 A:Accession: S62128
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-284 <KAW>
 A:Cross-references: EMBL:D38539; NID:g871835; PIDN:BAA07541.1; PID:g1208410
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994
 A:Accession: S58918
 A:Molecule type: protein
 A:Residues: 11-49; 171-211 <KAF>
 C:Genetics:
 A:Gene: tmy-1
 A:Introns: 15/3; 61/3; 80/3; 125/2; 164/3; 188/2; 234/3; 258/1
 C:Superfamily: tropomyosin
 C:Keywords: alternative splicing

Query Match 1.0%; Score 8; DB 2; Length 284;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 727 VEAQLKEA 734
 DB 144 VEAQLKEA 151
 |||||

RESULT 16
 T49361
 probable transport vesicle fusion protein SEC17 [imported] - Neurospora crassa
 N:Alternate names: protein Bld1.150
 C:Species: Neurospora crassa
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 A:Accession: T49361
 R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, R.
 submitted to the Protein Sequence Database, May 2000
 A:Reference number: Z25022
 A:Accession: T49361
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-292 <SCH>
 A:Cross-references: EMBL:AL355927; GSPDB:GN00116; NCSP:B1D1.150
 A:Experimental source: BAC clone Bld1; strain OR74A
 C:Genetics:
 A:Gene: NCSP:Bld1.150
 A:Map position: 6
 A:Introns: 11/3; 53/1; 155/2

A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916

A;Accession: H95115

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1039 <KUR>

A;Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:g14972477; GSPDB:GN00164; TIGR:SP4

A;Experimental source: strain TIGR4

C;Genetics:

A;Gene: SP1004

Query Match 4.0%; Score 32; DB 2; Length 1039;

Best Local Similarity 100.0%; Pred. No. 2.3e-23;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEOIVIKITDQGVVTSHTGHDHYHYNGKVPYDA 77

Db 67 AEOIVIKITDQGVVTSHTGHDHYHYNGKVPYDA 98

RESULT 8

D97985

hypothetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)

C;Species: Streptococcus pneumoniae

C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C;Accession: D97985

R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

y, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A;Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: D97985

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1039 <KUR>

A;Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:g15458515; GSPDB:GN00174

C;Genetics:

A;Gene: phtE

Query Match 4.0%; Score 32; DB 2; Length 1039;

Best Local Similarity 100.0%; Pred. No. 2.3e-23;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEOIVIKITDQGVVTSHTGHDHYHYNGKVPYDA 77

Db 67 AEOIVIKITDQGVVTSHTGHDHYHYNGKVPYDA 98

RESULT 9

T46758

hypothetical 92.4K protein - Streptococcus agalactiae

C;Species: Streptococcus agalactiae

C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000

C;Accession: T46758

R;Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heymann, J.; Schnitzler, N.; Lued

Infekt. Immun. 67, 871-878, 1999

A;Title: Lmb, a protein with similarities to the Irai adhesin family, mediates attachmen

A;Reference number: 224091; MUID:99115568; PMID:9916102

A;Accession: T46758

A;Status: preliminary;

A;Molecule type: DNA

A;Residues: 1-822 <SPE>

A;Cross-references: EMBL:AF062533; NID:g4249622; PIDN:AAD13797.1; PID:g4249624

A;Experimental source: strain R268

C;Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match

Best Local Similarity 100.0%; Pred. No. 3e-08;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISEELM 85

Db 92 YNGKVPYDAIISEELM 108

RESULT 10

F97985

hypothetical protein phtE-truncation [imported] - Streptococcus pneumoniae (strain R6)

C;Species: Streptococcus pneumoniae

C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C;Accession: F97985

R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A;Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: F97985

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-182 <KUR>

A;Cross-references: GB:AE007317; PIDN:AAK99714.1; PID:g15458517; GSPDB:GN00174

C;Genetics:

A;Gene: phtE-truncation

Query Match

Best Local Similarity 1.3%; Score 10; DB 2; Length 182;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 HGDHYHYIPK 197

Db 35 HGDHYHYIPK 44

RESULT 11

A48184

transcription initiation factor IID 110K chain - fruit fly (Drosophila melanogaster)

C;Species: Drosophila melanogaster

C;Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999

C;Accession: A48184

R;Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.

Proc. Natl. Acad. Sci. U.S.A. 90, 5896-5900, 1993

A;Title: The Drosophila 110-kDa transcription factor TFIIID subunit directly interacts wit

A;Reference number: A48184; MUID:93317591; PMID:8327460

A;Accession: A48184

A;Status: preliminary

A;Molecule type: mRNA; protein

A;Residues: 1-921 <KOK>

A;Cross-references: GB:S63550; NID:g398432; PID:g398433

A;Experimental source: embryo nuclear extract

A;Note: sequence extracted from NCBI backbone (NCBIN:134863, NCBIP:134864)

C;Genetics:

A;Gene: FlyBase:Taf110

A;Cross-references: FlyBase:FBgn0010280

C;Keywords: transcription initiation

Query Match

Best Local Similarity 1.1%; Score 9; DB 2; Length 921;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TMTSNNST 257

Db 178 TMTSNNST 186

RESULT 12

A45183

TBP-associated factor TFIIID - fruit fly (Drosophila sp.)

C;Species: Drosophila sp.

C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997

C;Accession: A45183

R;Hoey, T.; Weinzierl, R.O.; Gill, G.; Chen, J.L.; Dynlacht, B.D.; Tjian, R.

QY 740 KVTSSSLKANATETLAGLNNLTLOIMDNNSIMAEKILLALLKGSNPFSSVSKKIN 796
D98004
Db 772 KVTSSSLKANATETLAGLNNLTLOIMDNNSIMAEKILLALLKGSNPFSSVSKKIN 828

RESULT 3
D98004
histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98004
R:Hoskins, J.A.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Heaton, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Nelson, T.; Hickey, E.K.; Holt, I.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:gl5458682; GSPDB:GN00174
C:Genetics:
A:Gene: phpA
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 13.4%; Score 107; DB 2; Length 855;
Best Local Similarity 100.0%; Pred. No. 1.7e-99;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 KGGYVIVKDGKYYVYLKDAADNVRTKEENRQKQHSQHREGGTPRNDGAVALARSG 161
Db 133 KGGYVIVKDGKYYVYLKDAADNVRTKEENRQKQHSQHREGGTPRNDGAVALARSG 192

QY 162 RYTTDDGYIFNASDIETDGYIVPHGDHGHVYIPKNSLSASELAAA 208
Db 193 RYTTDDGYIFNASDIETDGYIVPHGDHGHVYIPKNSLSASELAAA 239

RESULT 4
B95136
conserved domain protein SP1174 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: B95136
R:Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heaton, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, S.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:gl4972654; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1174
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 7.5%; Score 60; DB 2; Length 819;
Best Local Similarity 100.0%; Pred. No. 7.3e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPEVSKREGINAEQIVIKITDQGYVTSHTGDHGHYNGKVPYDAIISELLMKDPNY 90
Db 52 ENLTPEVSKREGINAEQIVIKITDQGYVTSHTGDHGHYNGKVPYDAIISELLMKDPNY 111

RESULT 5
G95115
conserved hypothetical protein SP1003 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: G95115
R:Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heaton, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, S.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:gl4972476; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1003
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 7.5%; Score 60; DB 2; Length 839;
Best Local Similarity 100.0%; Pred. No. 7.5e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPEVSKREGINAEQIVIKITDQGYVTSHTGDHGHYNGKVPYDAIISELLMKDPNY 90
Db 52 ENLTPEVSKREGINAEQIVIKITDQGYVTSHTGDHGHYNGKVPYDAIISELLMKDPNY 111

RESULT 6
C97985
hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: C97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Heaton, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Nelson, T.; Hickey, E.K.; Holt, I.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: C97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:gl5458514; GSPDB:GN00174
C:Genetics:
A:Gene: phtD
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 7.5%; Score 60; DB 2; Length 853;
Best Local Similarity 100.0%; Pred. No. 7.6e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPEVSKREGINAEQIVIKITDQGYVTSHTGDHGHYNGKVPYDAIISELLMKDPNY 90
Db 52 ENLTPEVSKREGINAEQIVIKITDQGYVTSHTGDHGHYNGKVPYDAIISELLMKDPNY 111

RESULT 7
H95115
conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heaton, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, S.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappple, nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Lotius, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75284.1; PID:gl14972655; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1175
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 87.3%; Score 695; DB 2; Length 802;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRNVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 7 SYELGLYQARTVKNRNVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 66
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIVNEVKGYYIKVDGKYVYLKDA 120
Db 67 SHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIVNEVKGYYIKVDGKYVYLKDA 126
QY 121 AHADNVRTKEINRQKQHSQHREGGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 180
Db 127 AHADNVRTKEINRQKQHSQHREGGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 186
QY 181 GDVIVPHGDHYHYIPKNEISASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Db 187 GDVIVPHGDHYHYIPKNEISASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 246
QY 241 VSNPGTTNTNTSNNTNSQASQND:DSLKQYKLPLOSRHVESDGLVDPQAITSRT 300
Db 247 VSNPGTTNTNTSNNTNSQASQND:DSLKQYKLPLOSRHVESDGLVDPQAITSRT 306
QY 301 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHNWVPSRPEQSPQPTPEPSPG 360
Db 307 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHNWVPSRPEQSPQPTPEPSPG 366
QY 361 PQAPNLKIDNSLSVQLRVKVGEGYVFEKGLSRVVFAPKLPSETVKNLESKLSQES 420
Db 367 PQAPNLKIDNSLSVQLRVKVGEGYVFEKGLSRVVFAPKLPSETVKNLESKLSQES 426
QY 421 VSHLTAKENVAPRDQEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 480
Db 427 VSHLTAKENVAPRDQEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 486
QY 481 KEKLVDDLLAFAPITHPERLGPNSQIETDEVR:IAQLADKYTSDGYIFDEHDIISD 540
Db 487 KEKLVDDLLAFAPITHPERLGPNSQIETDEVR:IAQLADKYTSDGYIFDEHDIISD 546
QY 541 EGAAYVTPHGHSHWICKDSIDSKKVAQAQYTKGILPSPDADVKANPTGDSAAIY 600
Db 547 EGAAYVTPHGHSHWICKDSIDSKKVAQAQYTKGILPSPDADVKANPTGDSAAIY 606
QY 601 NRKVGKRIIPLVRLPYMVEHTEVKNGLNLIIPKDHYNHNIKFAWFDHDTYKAPNGYTTLED 660
Db 607 NRKVGKRIIPLVRLPYMVEHTEVKNGLNLIIPKDHYNHNIKFAWFDHDTYKAPNGYTTLED 666
QY 661 LFATIKYVVEHPDRPHNSDGNWASHV:LGKDHSEDPNKNFKADEPVEETPAEPEVP 720
Db 667 LFATIKYVVEHPDRPHNSDGNWASHV:LGKDHSEDPNKNFKADEPVEETPAEPEVP 726
QY 721 QVETEKVEAQLKEAEVLAKVTDSLSKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLA 780
Db 727 QVETEKVEAQLKEAEVLAKVTDSLSKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLA 786
QY 781 LLKGSNPSSVSKKIN 796

Db 787 LLKGSNPSSVSKKIN 802

RESULT 2

E98004
hypothetical protein phtA [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: E98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Ee
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, B.J.; Lu, J.; Matsushima, P.; McAhren, S.; Mc
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-828 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:gl15458683; GSPDB:GN00174
C:Genetics:
A:Gene: phtA
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 77.4%; Score 616; DB 2; Length 828;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 716; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 SEELMKDPNYKDKEDIVNEVKGYYIKVDGKYVYLKDAHADNVRTKEINRQKQEH 139
Db 112 SEELMKDPNYKDKEDIVNEVKGYYIKVDGKYVYLKDAHADNVRTKEINRQKQEH 171
QY 140 SQHREGTTPNDGAVALARSQGYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNE 199
Db 172 SQHREGTTPNDGAVALARSQGYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNE 231
QY 200 LSASELAAAFISGRGNLSNRTYRRQNSDNTSRTNWVPSNPGTTNTNTSNNTNS 259
Db 232 LSASELAAAFISGRGNLSNRTYRRQNSDNTSRTNWVPSNPGTTNTNTSNNTNS 291
QY 260 QASQNDIDSLKQYKLPLOSRHVESDGLVDPQAITSRTARGVAVPHGDHYHYIPYSQ 319
Db 292 QASQNDIDSLKQYKLPLOSRHVESDGLVDPQAITSRTARGVAVPHGDHYHYIPYSQ 351
QY 320 MSELERARIIPLYRSNHNWVPSRPEQSPQPTPEPSGPAPNLKIDNSLSVQL 379
Db 352 MSELERARIIPLYRSNHNWVPSRPEQSPQPTPEPSGPAPNLKIDNSLSVQL 411
QY 380 VRKVGEGYVFEKGLSRVVFAPKLPSETVKNLESKLSQESVSHLTAKENVAPRDQEF 439
Db 412 VRKVGEGYVFEKGLSRVVFAPKLPSETVKNLESKLSQESVSHLTAKENVAPRDQEF 471
QY 440 YDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTNKEKLVDDLLAFAPITHP 499
Db 472 YDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTNKEKLVDDLLAFAPITHP 531
QY 500 RLKGPSQIETDEVR:IAQLADKYTSDGYIFDEHDIISDEGDAYVTPHGHSHWICKD 559
Db 532 RLKGPSQIETDEVR:IAQLADKYTSDGYIFDEHDIISDEGDAYVTPHGHSHWICKD 591
QY 560 SLSDKKEVAAQAQYTKGILPSPDADVKANPTGDSAAAIYNRVKGEKRIPLVRLPYMVE 619
Db 592 SLSDKKEVAAQAQYTKGILPSPDADVKANPTGDSAAAIYNRVKGEKRIPLVRLPYMVE 651
QY 620 HTVEVKNGLNLIIPKDHYNHNIKFAWFDHDTYKAPNGYTTLEDLFTATIKYVVEHDEPHSN 679
Db 652 HTVEVKNGLNLIIPKDHYNHNIKFAWFDHDTYKAPNGYTTLEDLFTATIKYVVEHDEPHSN 711
QY 680 DGMGNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVPQVETEKVEAQLKEAEVL 739
Db 712 DGMGNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVPQVETEKVEAQLKEAEVL 771

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:20:59 ; Search time 27 Seconds
(without alignments)
2835.869 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SYELGLYQARTVKENRVS.....KLALLKGNPSSVSKEKIN 796

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database :

PIR_78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 695 | 87.3 | 802 | C95136 | conserved domain p |
| 2 | 616 | 77.4 | 828 | E98004 | hypothetical prote |
| 3 | 107 | 13.4 | 855 | D98004 | histidine Motif-Co |
| 4 | 60 | 7.5 | 819 | B95136 | conserved domain p |
| 5 | 60 | 7.5 | 839 | G95115 | conserved hypothet |
| 6 | 60 | 7.5 | 853 | C97985 | hypothetical prote |
| 7 | 32 | 4.0 | 1039 | H95115 | conserved hypothet |
| 8 | 32 | 4.0 | 1039 | D97985 | hypothetical prote |
| 9 | 17 | 2.1 | 822 | T46758 | hypothetical 92.4K |
| 10 | 10 | 1.3 | 182 | F97985 | hypothetical prote |
| 11 | 9 | 1.1 | 921 | A48184 | transcription init |
| 12 | 9 | 1.1 | 921 | A45183 | TBP-associated fac |
| 13 | 8 | 1.0 | 71 | E72853 | AcOrf-29 protein - |
| 14 | 8 | 1.0 | 284 | S58921 | tropomyosin isofo |
| 15 | 8 | 1.0 | 284 | S58922 | tropomyosin isofo |
| 16 | 8 | 1.0 | 292 | T49361 | probable transport |
| 17 | 8 | 1.0 | 294 | T27791 | hypothetical prote |
| 18 | 8 | 1.0 | 448 | S05355 | hypothetical prote |
| 19 | 8 | 1.0 | 484 | F82203 | probable multidrug |
| 20 | 8 | 1.0 | 488 | B89957 | hypothetical prote |
| 21 | 8 | 1.0 | 795 | S73830 | endopeptidase Ia (|
| 22 | 8 | 1.0 | 825 | T00818 | hypothetical prote |
| 23 | 8 | 1.0 | 840 | C75313 | probable proteinas |
| 24 | 8 | 1.0 | 945 | A64714 | helicase - Helicob |
| 25 | 8 | 1.0 | 1289 | B70748 | probable smc prote |
| 26 | 8 | 1.0 | 1659 | H97926 | hypothetical prote |
| 27 | 8 | 1.0 | 1659 | G95057 | endo-beta-N-acetyl |
| 28 | 7 | 0.9 | 44 | S55073 | env polyprotein - |
| 29 | 7 | 0.9 | 49 | S55072 | env polyprotein - |

RESULT 1

C95136 conserved domain protein SP1175 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95136
R:Tetreltin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

ALIGNMENTS

| | | | | | | |
|----|---|-----|-----|---|--------|--------------------|
| 30 | 7 | 0.9 | 49 | 2 | D70091 | hypothetical prote |
| 31 | 7 | 0.9 | 68 | 2 | D87657 | hypothetical prote |
| 32 | 7 | 0.9 | 70 | 2 | A75134 | hypothetical prote |
| 33 | 7 | 0.9 | 90 | 2 | C90404 | conserved hypothet |
| 34 | 7 | 0.9 | 93 | 2 | C75396 | hypothetical prote |
| 35 | 7 | 0.9 | 94 | 2 | E75556 | hypothetical prote |
| 36 | 7 | 0.9 | 100 | 1 | BVNYBA | chaperonin groES - |
| 37 | 7 | 0.9 | 100 | 1 | BVNY7B | chaperonin groES - |
| 38 | 7 | 0.9 | 101 | 2 | T38824 | thioltransferase - |
| 39 | 7 | 0.9 | 101 | 2 | E82874 | ribosomal protein |
| 40 | 7 | 0.9 | 106 | 2 | I64171 | hypothetical prote |
| 41 | 7 | 0.9 | 110 | 2 | D87498 | HesB/yadR/lfhF fam |
| 42 | 7 | 0.9 | 110 | 2 | C86891 | hypothetical prote |
| 43 | 7 | 0.9 | 117 | 2 | AG3598 | transcription regu |
| 44 | 7 | 0.9 | 121 | 2 | T44473 | conserved hypothet |
| 45 | 7 | 0.9 | 124 | 1 | NRPRH | pancreatic ribonuc |
| 46 | 7 | 0.9 | 128 | 2 | H72500 | hypothetical prote |
| 47 | 7 | 0.9 | 129 | 2 | T36220 | mutr-like protein |
| 48 | 7 | 0.9 | 130 | 2 | S76934 | hypothetical prote |
| 49 | 7 | 0.9 | 131 | 2 | T35461 | mutr-like protein |
| 50 | 7 | 0.9 | 136 | 2 | B65240 | hypothetical prote |
| 51 | 7 | 0.9 | 136 | 2 | D64964 | hypothetical prote |
| 52 | 7 | 0.9 | 136 | 2 | T19240 | hypothetical prote |
| 53 | 7 | 0.9 | 139 | 2 | B81286 | probable membrane |
| 54 | 7 | 0.9 | 143 | 2 | A71717 | hypothetical prote |
| 55 | 7 | 0.9 | 145 | 2 | AF1219 | Salmonella enteric |
| 56 | 7 | 0.9 | 147 | 2 | D83261 | probable transcrip |
| 57 | 7 | 0.9 | 148 | 2 | A86079 | PTS system, fructo |
| 58 | 7 | 0.9 | 148 | 2 | E91232 | PTS system, fructo |
| 59 | 7 | 0.9 | 148 | 2 | D48649 | frva protein - Esc |
| 60 | 7 | 0.9 | 154 | 2 | S39873 | hypothetical prote |
| 61 | 7 | 0.9 | 157 | 2 | AC2947 | ribonuclease H [im |
| 62 | 7 | 0.9 | 157 | 2 | G98135 | ribonuclease H PAI |
| 63 | 7 | 0.9 | 157 | 2 | AF3240 | conserved hypothet |
| 64 | 7 | 0.9 | 158 | 1 | D63256 | hypothetical prote |
| 65 | 7 | 0.9 | 158 | 2 | A89943 | transcription elon |
| 66 | 7 | 0.9 | 165 | 2 | C83415 | hypothetical prote |
| 67 | 7 | 0.9 | 166 | 2 | A86450 | probable glycine c |
| 68 | 7 | 0.9 | 178 | 2 | C81235 | transcription anti |
| 69 | 7 | 0.9 | 179 | 2 | H71920 | adenine phosphorib |
| 70 | 7 | 0.9 | 181 | 2 | A70178 | conserved hypothet |
| 71 | 7 | 0.9 | 185 | 2 | F75487 | v-type ATP synthas |
| 72 | 7 | 0.9 | 191 | 2 | T35792 | beta-glucosidase - |
| 73 | 7 | 0.9 | 197 | 1 | S01249 | cytochrome c-L pre |
| 74 | 7 | 0.9 | 201 | 2 | S06147 | GTP-binding protei |
| 75 | 7 | 0.9 | 203 | 2 | F90658 | hypothetical prote |
| 76 | 7 | 0.9 | 203 | 2 | F85509 | hypothetical prote |
| 77 | 7 | 0.9 | 210 | 2 | S62443 | hypothetical prote |
| 78 | 7 | 0.9 | 211 | 2 | AH0899 | diadenosine 5', 5' |
| 79 | 7 | 0.9 | 211 | 2 | T01627 | conserved hypothet |
| 80 | 7 | 0.9 | 213 | 2 | F83144 | probable ATP bindi |
| 81 | 7 | 0.9 | 219 | 2 | B75533 | conserved hypothet |
| 82 | 7 | 0.9 | 226 | 2 | D65105 | hypothetical prote |
| 83 | 7 | 0.9 | 226 | 2 | A98133 | hypothetical prote |
| 84 | 7 | 0.9 | 226 | 2 | D85978 | hypothetical prote |
| 85 | 7 | 0.9 | 226 | 2 | G81894 | probable integral |
| 86 | 7 | 0.9 | 232 | 2 | D81125 | hypothetical prote |
| 87 | 7 | 0.9 | 232 | 2 | C83744 | two-component resp |
| 88 | 7 | 0.9 | 237 | 2 | G87286 | conserved hypothet |
| 89 | 7 | 0.9 | 238 | 2 | I67638 | proteasome activat |
| 90 | 7 | 0.9 | 240 | 2 | D75203 | hypothetical prote |

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 90
 |||||
 Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 92

RESULT 24

AAB12763
 ID AAB12763 standard; protein; 811 AA.
 XX
 AC AAB12763;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain RX1 BVH-11 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otis media, bacteraemia and/or pneumonia.
 XX
 PS Disclosure; Fig 12; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
 CC antigen, from the present invention

XX SQ Sequence 811 AA;

Query Match 7.5%; Score 60; DB 3; Length 811;
 Best Local Similarity 100.0%; Pred. No. 2e-48;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 90
 |||||
 Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 92

RESULT 25

AAB12758
 ID AAB12758 standard; protein; 816 AA.
 XX
 AC AAB12758;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain WU2 BVH-11-2 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 XX otitis media; pneumonia; immunisation; bactericidal.
 OS Streptococcus pneumoniae.

XX WO200039299-A2.

XX PD 06-JUL-2000.

XX PF 20-DEC-1999; 99WO-CA001218.

XX PR 23-DEC-1998; 98US-0113800P.

XX PA (BIOC-) BIOCHEM PHARMA INC.

XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX DR WPI; 2000-452397/39.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otis media, bacteraemia and/or pneumonia.

XX PS Disclosure; Fig 12; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11-2
 CC protein antigen, from the present invention

XX SQ Sequence 816 AA;

Query Match 7.5%; Score 60; DB 3; Length 816;
 Best Local Similarity 100.0%; Pred. No. 2e-48;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 90

Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 92

Search completed: October 1, 2004, 07:32:24
 Job time : 84 secs

PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA001218.
XX
XX 23-DEC-1998; 98US-0113800P.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
DR
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX
PS Disclosure; Fig 12; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
CC antigen, from the present invention
XX
XX Sequence 811 AA;
SQ

Query Match 7.5%; Score 60; DB 3; Length 811;
Best Local Similarity 100.0%; Pred. No. 2e-48; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 0;
QY 31 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 90
DB 33 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 92

RESULT 22
AAB12762
ID AAB12762 standard; protein; 811 AA.
XX
AC AAB12762;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain A66 BVH-11 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA001218.
XX
XX 23-DEC-1998; 98US-0113800P.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
DR
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX
PS Disclosure; Fig 12; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
CC antigen, from the present invention
XX
XX Sequence 811 AA;
SQ

Query Match 7.5%; Score 60; DB 3; Length 811;
Best Local Similarity 100.0%; Pred. No. 2e-48; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 0;
QY 31 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 90
DB 33 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 92

RESULT 22
AAB12762
ID AAB12762 standard; protein; 811 AA.
XX
AC AAB12762;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain A66 BVH-11 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA001218.
XX
XX 23-DEC-1998; 98US-0113800P.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
DR
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX

PS Disclosure; Fig 12; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
CC antigen, from the present invention
XX
XX Sequence 811 AA;
SQ

Query Match 7.5%; Score 60; DB 3; Length 811;
Best Local Similarity 100.0%; Pred. No. 2e-48; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 0;
QY 31 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 90
DB 33 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 92

RESULT 23
AAB12760
ID AAB12760 standard; protein; 811 AA.
XX
AC AAB12760;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain P4241 BVH-11 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA001218.
XX
XX 23-DEC-1998; 98US-0113800P.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
DR
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX
PS Disclosure; Fig 12; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
CC antigen, from the present invention
XX
XX Sequence 811 AA;
SQ

Query Match 7.5%; Score 60; DB 3; Length 811;
Best Local Similarity 100.0%; Pred. No. 2e-48; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 0;
QY 31 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 90
DB 33 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 92

RESULT 23
AAB12760
ID AAB12760 standard; protein; 811 AA.
XX
AC AAB12760;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain P4241 BVH-11 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA001218.
XX
XX 23-DEC-1998; 98US-0113800P.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
DR
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX
PS Disclosure; Fig 12; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
CC antigen, from the present invention
XX
XX Sequence 811 AA;
SQ

Query Match 7.5%; Score 60; DB 3; Length 811;
Best Local Similarity 100.0%; Pred. No. 2e-48; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 0;
QY 31 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 90
DB 33 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 92

RESULT 23
AAB12760
ID AAB12760 standard; protein; 811 AA.
XX
AC AAB12760;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain P4241 BVH-11 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA001218.
XX
XX 23-DEC-1998; 98US-0113800P.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
DR
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX

CC represented in the printed specification, but is based on sequence
 CC information supplied by the European Patent Office
 XX
 SQ Sequence 840 AA;

Query Match 10.1%; Score 80; DB 7; Length 840;
 Best Local Similarity 100.0%; Pred. No. 1e-67;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 70
 |||||
 Db 31 TVKENRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 90
 |||||
 QY 71 GKVPYDAIISELLMKDPNY 90
 |||||
 Db 91 GKVPYDAIISELLMKDPNY 110
 |||||

RESULT 19
 AAB12764
 ID AAB12764 standard; protein; 805 AA.
 XX
 AC AAB12764;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain JNR7/87 BVH-11 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia.
 PS Disclosure; Fig 12; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
 CC antigen, from the present invention
 XX
 SQ Sequence 805 AA;

Query Match 7.5%; Score 60; DB 3; Length 805;
 Best Local Similarity 100.0%; Pred. No. 2e-48;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNKVPYDAIISELLMKDPNY 90
 |||||
 Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNKVPYDAIISELLMKDPNY 92
 |||||

RESULT 21
 AAB12761
 ID AAB12761 standard; protein; 811 AA.
 XX
 AC AAB12761;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain WU2 BVH-11 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia.
 PS Disclosure; Fig 12; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
 CC antigen, from the present invention
 XX
 SQ Sequence 805 AA;

RESULT 20
 AAB12765
 ID AAB12765 standard; protein; 807 AA.
 XX
 AC AAB12765;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain SP63 BVH-11 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia.
 PS Disclosure; Fig 12; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
 CC antigen, from the present invention
 XX
 SQ Sequence 807 AA;

Query Match 7.5%; Score 60; DB 3; Length 807;
 Best Local Similarity 100.0%; Pred. No. 2e-48;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNKVPYDAIISELLMKDPNY 90
 |||||
 Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNKVPYDAIISELLMKDPNY 92
 |||||

RESULT 21
 AAB12761
 ID AAB12761 standard; protein; 811 AA.
 XX
 AC AAB12761;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain WU2 BVH-11 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia.
 PS Disclosure; Fig 12; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
 CC antigen, from the present invention
 XX
 SQ Sequence 807 AA;

XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens, have
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence represents the S. pneumoniae BVH-11
CC protein antigen
XX
SQ Sequence 840 AA;
Query Match 10.1%; Score 80; DB 3; Length 840;
Best Local Similarity 100.0%; Pred. No. 1e-67;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 TVKENRVSYIDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVTSRHDHYHYN 70
DB 31 TVKENRVSYIDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVTSRHDHYHYN 90
QY 71 GKVPYDAIISBELLMKDPNY 90
DB 91 GKVPYDAIISBELLMKDPNY 110
RESULT 17
AAU75933
ID AAU75933 standard; protein; 840 AA.
XX
AC AAU75933;
XX
DT 08-MAY-2002 (first entry)
XX
DE Streptococcus pneumoniae BVH-11 protein.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
KW streptococcal bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA000909.
XX
PR 20-JUN-2000; 2000US-0212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
DR WPI; 2002-122272/16.
XX
DR N-PSDB; ABK15103.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
PT polypeptides, useful as vaccine components for treating or preventing
PT streptococcal infections such as otitis media, meningitis, and
PT bacteraemia.
XX
PS Example 1; Fig 7; 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90% identity
CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or

CC Staphylococcus aureus) in an individual susceptible to the infection. A
CC polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This is the amino acid sequence of
CC Streptococcus pneumoniae protein BVH-11, used to create the antigenic
CC peptides described in the method of the invention
XX
SQ Sequence 840 AA;
Query Match 10.1%; Score 80; DB 5; Length 840;
Best Local Similarity 100.0%; Pred. No. 1e-67;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 TVKENRVSYIDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVTSRHDHYHYN 70
DB 31 TVKENRVSYIDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVTSRHDHYHYN 90
QY 71 GKVPYDAIISBELLMKDPNY 90
DB 91 GKVPYDAIISBELLMKDPNY 110
RESULT 18
ABM18797
ID ABM18797 standard; protein; 840 AA.
XX
AC ABM18797;
XX
DT 13-OCT-2003 (first entry)
XX
DE S. pneumoniae BVH-11 polypeptide SEQ ID NO: 8.
XX
KW antibacterial; antiinflammatory; auditory; vaccine; meningitis;
KW streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3.
XX
OS Streptococcus pneumoniae.
XX
PN WO2003054007-A2.
XX
PD 03-JUL-2003.
XX
PF 20-DEC-2002; 2002WO-CA002006.
XX
PR 20-DEC-2001; 2001US-0341252P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;
XX
DR Labbe S;
XX
DR WPI; 2003-569224/53.
XX
PT New isolated polypeptides of Streptococcus pneumoniae, useful for
PT diagnosing, preventing or treating streptococcal infection, meningitis,
PT otitis media, bacteraemia or pneumonia infection.
XX
PS Example 1; Fig 10; 79pp; English.
XX
CC The invention relates to a novel isolated polypeptide of Streptococcus
CC pneumoniae. A polypeptide of the invention has antibacterial,
CC antiinflammatory, and auditory activity, and is used as a vaccine. The
CC polypeptide or pharmaceutical composition is useful for the prophylactic
CC or therapeutic treatment of streptococcal infection, meningitis, otitis
CC media, bacteraemia or pneumonia infection. The kit is useful for
CC detecting or diagnosing streptococcal infection. The pharmaceutical
CC composition is useful as a vaccine. The polynucleotides are useful in
CC designing DNA probes for detecting circulating Streptococcus in a
CC biological sample. The present sequence is used in the exemplification of
CC the invention. Note: The sequence data for this patent is not fully


```

XX 21-NOV-2000 (first entry)
DT Streptococcus pneumoniae strain SP64 BVH-11 protein antigen.
DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX Streptococcus pneumoniae.
OS WO200039299-A2.
XX 06-JUL-2000.
PD 20-DEC-1999; 99WO-CA001218.
XX 23-DEC-1998; 98US-0113800P.
PF (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX Disclosure; Fig 12; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein
CC antigen, from the present invention
XX Sequence 821 AA;
SQ
Query Match 10.1%; Score 80; DB 3; Length 821;
Best Local Similarity 100.0%; Pred. No. 1e-67;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 TVKENNRVSYIDGKQATKTNLTDPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 70
Db 12 TVKENNRVSYIDGKQATKTNLTDPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 71
QY 71 GKVPYDAIISSELLMKDPNY 90
Db 72 GKVPYDAIISSELLMKDPNY 91
RESULT 13
AAB12727
ID AAB12727 standard; protein; 821 AA.
XX AAB12727;
XX 21-NOV-2000 (first entry)
DE Streptococcus pneumoniae BVH-11M protein antigen SEQ ID NO:60.
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX Streptococcus pneumoniae.
OS WO200039299-A2.
XX

```

```

PD 06-JUL-2000.
XX 20-DEC-1999; 99WO-CA001218.
XX 23-DEC-1998; 98US-0113800P.
PF (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX Claim 18; Fig 25; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence represents the S. pneumoniae BVH-11M
CC protein antigen
XX Sequence 821 AA;
SQ
Query Match 10.1%; Score 80; DB 3; Length 821;
Best Local Similarity 100.0%; Pred. No. 1e-67;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 TVKENNRVSYIDGKQATKTNLTDPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 70
Db 12 TVKENNRVSYIDGKQATKTNLTDPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 71
QY 71 GKVPYDAIISSELLMKDPNY 90
Db 72 GKVPYDAIISSELLMKDPNY 91
RESULT 14
AAU84026
ID AAU84026 standard; peptide; 821 AA.
XX AAU84026;
XX 08-MAY-2002 (first entry)
DT Truncated variant of S. pneumoniae BVH-11, BVH-11M.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
KW streptococcal bacterial infection; mutant; mutain.
XX Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
XX 27-DEC-2001.
XX 19-JUN-2001; 2001WO-CA000908.
XX 20-JUN-2000; 2000US-0212683P.
PR (SHIR-) SHIRE BIOCHEM INC.
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing

```

RESULT 10
 AAU84093
 ID AAU84093 standard; peptide; 690 AA.
 XX
 AC AAU84093;
 DT 08-MAY-2002 (first entry)
 XX
 DE Truncated variant of S. pneumoniae BVH-11, NEW16.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
 KW streptococcal bacterial infection; mutant; muten.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA000908.
 XX
 PR 20-JUN-2000; 2000US-0212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 DR WPI; 2002-12272/16.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
 PT polypeptides, useful as vaccine components for treating or preventing
 PT streptococcal infections such as otitis media, meningitis, and
 PT bacteremia.
 XX
 PS Example 1; Page; 113pp; English.
 CC
 CC The invention describes an isolated polypeptide (I) with 70-90% identity
 CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
 CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
 CC Staphylococcus aureus) in an individual susceptible to the infection. A
 CC polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention. Note: This sequence does not
 CC appear in the specification but has been created according to information
 CC given in the invention
 XX
 SQ Sequence 690 AA;
 Query Match 10.1%; Score 80; DB 5; Length 690;
 Best Local Similarity 100.0%; Pred. No. 8.8e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 TVKENNRVSYIDGKQATKNTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70
 DB 12 TVKENNRVSYIDGKQATKNTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71
 QY 71 GKVPYDAIISELLMKDPNY 90
 DB 72 GKVPYDAIISELLMKDPNY 91

RESULT 11
 ABM18826
 ID ABM18826 standard; protein; 690 AA.
 XX
 AC ABM18826;
 DT 13-OCT-2003 (first entry)
 XX
 DE S. pneumoniae variant protein NEW16.
 XX
 KW antibacterial; antiinflammatory; auditory; vaccine; meningitis;
 KW streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3;
 KW mutant; muten; New 43; BVH-11; BVH-11-2.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX
 PN WO2003054007-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-CA002006.
 XX
 PR 20-DEC-2001; 2001US-0341252P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;
 PI Labbe S;
 XX
 DR WPI; 2003-569224/53.
 XX
 PT New isolated polypeptides of Streptococcus pneumoniae, useful for
 PT diagnosing, preventing or treating streptococcal infection, meningitis,
 PT otitis media, bacteraemia or pneumonia infection.
 XX
 PS Example 1; SEQ ID NO 38; 79pp; English.
 CC
 CC The invention relates to a novel isolated polypeptide of Streptococcus
 CC pneumoniae. A polypeptide of the invention has antibacterial,
 CC antiinflammatory, and auditory activity, and is used as a vaccine. The
 CC polypeptide or pharmaceutical composition is useful for the prophylactic
 CC or therapeutic treatment of streptococcal infection, meningitis, otitis
 CC media, bacteraemia or pneumonia infection. The kit is useful for
 CC detecting or diagnosing streptococcal infection. The pharmaceutical
 CC composition is useful as a vaccine. The polynucleotides are useful in
 CC designing DNA probes for detecting circulating Streptococcus in a
 CC biological sample. The present sequence is used in the exemplification of
 CC the invention. Note: The sequence data for this patent is not fully
 CC represented in the printed specification, but is based on sequence
 CC information supplied by the European Patent Office
 XX
 SQ Sequence 690 AA;
 Query Match 10.1%; Score 80; DB 7; Length 690;
 Best Local Similarity 100.0%; Pred. No. 8.8e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 TVKENNRVSYIDGKQATKNTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70
 DB 12 TVKENNRVSYIDGKQATKNTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71
 QY 71 GKVPYDAIISELLMKDPNY 90
 DB 72 GKVPYDAIISELLMKDPNY 91

RESULT 12
 AAB12766
 ID AAB12766 standard; protein; 821 AA.
 XX
 AC AAB12766;

CC described in the method of the invention. Note: This sequence does not
 CC appear in the specification but has been created according to information
 CC given in the invention

XX SQ Sequence 334 AA;

Query Match 10.1%; Score 80; DB 5; Length 334;
 Best Local Similarity 100.0%; Pred. No. 4.5e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70
 |||
 Db 12 TVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71

QY 71 GKVPYDAIISELLMKDPNY 90
 |||
 Db 72 GKVPYDAIISELLMKDPNY 91

RESULT 8

ABM18808
 ID ABM18808 standard; protein; 334 AA.

XX AC ABM18808;

DT 13-OCT-2003 (first entry)

XX DE S. pneumoniae variant protein BVH-11A.

XX KW antibacterial; antiinflammatory; auditory; vaccine; meningitis;
 KW streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3;
 KW mutant; mutein; New 43; BVH-11; BVH-11-2.

XX OS Streptococcus pneumoniae.
 OS Synthetic.

PN WO2003054007-A2.

XX PD 03-JUL-2003.

XX PF 20-DEC-2002; 2002WO-CA002006.

XX PR 20-DEC-2001; 2001US-0341252P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;
 PI Labbe S;

XX DR WPI; 2003-569224/53.

XX PT New isolated polypeptides of Streptococcus pneumoniae, useful for
 PT diagnosing, preventing or treating streptococcal infection, meningitis,
 PT otitis media, bacteraemia or pneumonia infection.

XX PS Example 1; SEQ ID NO 20; 79pp; English.

XX CC The invention relates to a novel isolated polypeptide of Streptococcus
 CC pneumoniae. A polypeptide of the invention has antibacterial,
 CC antiinflammatory, and auditory activity, and is used as a vaccine. The
 CC polypeptide or pharmaceutical composition is useful for the prophylactic
 CC or therapeutic treatment of streptococcal infection, meningitis, otitis
 CC media, bacteraemia or pneumonia infection. The kit is useful for
 CC detecting or diagnosing streptococcal infection. The pharmaceutical
 CC composition is useful as a vaccine. The polynucleotides are useful in
 CC designing DNA probes for detecting circulating Streptococcus in a
 CC biological sample. The present sequence is used in the exemplification of
 CC the invention. Note: The sequence data for this patent is not fully
 CC represented in the printed specification, but is based on sequence
 CC information supplied by the European Patent Office

XX SQ Sequence 334 AA;

Query Match 10.1%; Score 80; DB 7; Length 334;
 Best Local Similarity 100.0%; Pred. No. 4.5e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70
 |||
 Db 12 TVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71

QY 71 GKVPYDAIISELLMKDPNY 90
 |||
 Db 72 GKVPYDAIISELLMKDPNY 91

RESULT 9

AAB12745
 ID AAB12745 standard; protein; 690 AA.

XX AC AAB12745;

XX DT 23-NOV-2000 (first entry)

XX DE Streptococcus pneumoniae NEW16 protein antigen SEQ ID NO:79.

XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.

XX OS Streptococcus pneumoniae.

XX PN WO200039299-A2.

XX PD 06-JUL-2000.

XX PF 20-DEC-1999; 99WO-CA001218.

XX PR 23-DEC-1998; 98US-0113800P.

XX PA (BIOC-) BIOCHEM PHARMA INC.

XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX DR WPI; 2000-452397/39.

XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia.

XX PS Claim 18; Fig 44; 106pp; English.

XX CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence represents the S. pneumoniae NEW16
 CC protein antigen

XX SQ Sequence 690 AA;

Query Match 10.1%; Score 80; DB 3; Length 690;
 Best Local Similarity 100.0%; Pred. No. 8.8e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70
 |||
 Db 12 TVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71

QY 71 GKVPYDAIISELLMKDPNY 90
 |||
 Db 72 GKVPYDAIISELLMKDPNY 91

Qy 421 VSHTLTAKKNVAPDQDFYKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
 Db 441 VSHTLTAKKNVAPDQDFYKAYNLLTEAHKALFENKGRNSDFQALDKLLERLNDESTN 500
 Qy 481 KEKLVDDLLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540
 Db 501 KEKLVDDLLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 560
 Qy 541 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAQYTKKGLPPSPADVKANPTGDSAAAIY 600
 Db 561 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAQYTKKGLPPSPADVKANPTGDSAAAIY 620
 Qy 601 NRKVGKRIPLVRLPYMVEHTVEVKNGLIIPKHDYHNKIFAFDDHTYKAPNGYTTLED 660
 Db 621 NRKVGKRIPLVRLPYMVEHTVEVKNGLIIPKHDYHNKIFAFDDHTYKAPNGYTTLED 680
 Qy 661 LFATIKYVVEHPDRPHNSDGNWGNASBEHLGKDHSEDPNKNFKADEEPVEETPAEPEVP 720
 Db 681 LFATIKYVVEHPDRPHNSDGNWGNASBEHLGKDHSEDPNKNFKADEEPVEETPAEPEVP 740
 Qy 721 QVTEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLA 780
 Db 741 QVTEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLA 800
 Qy 781 LLKGSNPSSVSKEKIN 796
 Db 801 LLKGSNPSSVSKEKIN 816

RESULT 6
 AAB12728
 ID AAB12728 standard; protein; 334 AA.
 AC AAB12728;
 XX
 DT 21-NOV-2000 (first entry)
 DE Streptococcus pneumoniae BVH-11A protein antigen SEQ ID NO:61.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI; 2000-452397/39.
 DR
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otis media, bacteraemia and/or pneumonia.
 XX
 PS Claim 18; Fig 26; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence represents the S. pneumoniae BVH-11A

CC protein antigen
 XX Sequence 334 AA;
 SQ
 Query Match 10.1%; Score 80; DB 3; Length 334;
 Best Local Similarity 100.0%; Pred. No. 4.5e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 11 TVKENRVSYDVGKQATKQKTENLTTPDEVSKREGINAEQIVIKITDQGYVTSHGSHYHYN 70
 Db 12 TVKENRVSYDVGKQATKQKTENLTTPDEVSKREGINAEQIVIKITDQGYVTSHGSHYHYN 71
 Qy 71 GKVPYDAIISEELMKDPNY 90
 Db 72 GKVPYDAIISEELMKDPNY 91

RESULT 7
 AAU84027
 ID AAU84027 standard; peptide; 334 AA.
 AC AAU84027;
 XX
 DT 08-MAY-2002 (first entry)
 DE Truncated variant of S. pneumoniae BVH-11, BVH-11A.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
 KW streptococcal bacterial infection; mutant; mutain.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA000908.
 XX
 PR 20-JUN-2000; 2000US-0212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 DR
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
 PT polypeptides, useful as vaccine components for treating or preventing
 PT streptococcal infections such as otitis media, meningitis, and
 PT bacteraemia.
 XX
 PS Example 1; Page; 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90% identity
 CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
 CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection. A
 CC polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

241 VSNPGTTNTSNTNSQASQNDISLLKQLYKPLSQHVSDGLVDPDAITST 300
 247 VSNPGTTNTSNTNSQASQNDISLLKQLYKPLSQHVSDGLVDPDAITST 306
 301 ARGVAVPHGDHYHFIPIYSQMSLEERIIARIIPLYRSNHWVDSRPEQSPQTPPEPSG 360
 307 ARGVAVPHGDHYHFIPIYSQMSLEERIIARIIPLYRSNHWVDSRPEQSPQTPPEPSG 366
 361 POPAPNLKIDNSLSVQLVRKVGEGYVPEEKIGISRYVFAKDLPSKLVESKLSKQES 420
 367 POPAPNLKIDNSLSVQLVRKVGEGYVPEEKIGISRYVFAKDLPSKLVESKLSKQES 426
 421 VSHLTAKENAVAPRQDFYDQYKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 480
 427 VSHLTAKENAVAPRQDFYDQYKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 486
 481 KEKLVDDLLAFIAPITHPERLGKPNQIETYTEDEVRIALADKYTTSDGYIFDEHDIISD 540
 487 KEKLVDDLLAFIAPITHPERLGKPNQIETYTEDEVRIALADKYTTSDGYIFDEHDIISD 546
 541 EGDVYVTHMGHSHWIGKDSLSDEKVAQAAYTKEKGIILPPSPDADVKANPTGDSAAAY 600
 547 EGDVYVTHMGHSHWIGKDSLSDEKVAQAAYTKEKGIILPPSPDADVKANPTGDSAAAY 606
 601 NRKVGKRIPLVRLPMVEHTVEVKNGLNLIIPHDKDHYHNIKFAWFDHTYKAPNGYTLED 660
 607 NRKVGKRIPLVRLPMVEHTVEVKNGLNLIIPHDKDHYHNIKFAWFDHTYKAPNGYTLED 666
 661 LPATIKYVVEHDPHSDNGMGNASEHVLGKKHSDENKKNFKADEEVEETPAEPPV 720
 667 LPATIKYVVEHDPHSDNGMGNASEHVLGKKHSDENKKNFKADEEVEETPAEPPV 726
 721 QVETEKVEAQLKAEVLLAKVTDSSKANATETIAGLRNLTLOIMDNNSINABAEKLLA 780
 727 QVETEKVEAQLKAEVLLAKVTDSSKANATETIAGLRNLTLOIMDNNSINABAEKLLA 786
 781 LKGSNPSSVSKEKIN 796
 787 LKGSNPSSVSKEKIN 802

RESULT 5
 AAB01468
 ID AAB01468 standard; protein; 819 AA.
 XX AAB01468;
 AC AAB01468;
 20-OCT-2000 (first entry)
 XX Recombinant variant of Sp36 (Sp36A) of *S. pneumoniae*.
 DE Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia.
 XX Streptococcus pneumoniae.
 OS

Key Location/Qualifiers
 FT Region 63...68
 FT /label= Histidine triad residue
 FT Region 118...145
 FT /label= Coiled coil region
 FT Region 189...194
 FT /label= Histidine triad residue
 FT Region 309...314
 FT /label= Histidine triad residue
 FT Region 406...434
 FT /label= Coiled coil region
 FT Region 462...493
 FT /label= Coiled coil region
 FT Region 550...555
 FT /label= Histidine triad residue

634...639
 /label= Histidine triad residue
 724...751
 /label= Coiled coil region

WO200037105-A2.
 29-JUN-2000.
 21-DEC-1999; 99WO-US030390.
 21-DEC-1998; 98US-0113048P.
 (MEDI-) MEDIMUNE INC.
 Johnson LS, Koenig S, Adamou JE;
 WPI; 2000-452129/39.
 DR N-PSDB; AAA47604.

Vaccine useful for prophylaxis and treatment of pneumococcal infections
 such as otitis media, nasopharyngeal and bronchial infections, comprises
 Streptococcus pneumoniae proteins.

Claim 1; Page 61-64; 70pp; English.

Although a number of proteins have been suggested as being involved in
 the pathogenicity of Streptococcus pneumoniae, there still remains a need
 to identify polypeptides having epitopes in common from various strains
 of *S. pneumoniae* in order to utilize such polypeptides in vaccines to
 protect against a wide variety of *S. pneumoniae*. New vaccine compositions
 are described which comprise a Streptococcus pneumoniae polypeptide (or
 fragments) of 80 - 680 amino acids in length that comprise at least one
 histidine triad residue (HxxHxH) or a coiled-coil region, or an antibody
 directed against these features. The vaccine is useful in protecting
 against infection by Streptococcus pneumoniae. The vaccine composition
 comprising antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal and
 CC bronchial infections

XX Sequence 819 AA;
 SQ

Query Match 87.3%; Score 695; DB 3; Length 819;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
 DB 21 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 80
 QY 61 SHGDHYHYNGKVPYDAIISEELLMKDPNYKLDKDEDIVNEVKGYYIVKDGKYYVYLKDA 120
 DB 81 SHGDHYHYNGKVPYDAIISEELLMKDPNYKLDKDEDIVNEVKGYYIVKDGKYYVYLKDA 140
 QY 121 AHADNVRTKEEINRQKQESHQHREGTTPNDGAVALARSGRYTTDDGYIFNADIIEDT 180
 DB 141 AHADNVRTKEEINRQKQESHQHREGTTPNDGAVALARSGRYTTDDGYIFNADIIEDT 200
 QY 181 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRYRQNSNTSRTNTWVPS 240
 DB 201 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRYRQNSNTSRTNTWVPS 260
 QY 241 VSNPGTTNTSNTNSQASQNDISLLKQLYKPLSQHVSDGLVDPDAITST 300
 DB 261 VSNPGTTNTSNTNSQASQNDISLLKQLYKPLSQHVSDGLVDPDAITST 320
 QY 301 ARGVAVPHGDHYHFIPIYSQMSLEERIIARIIPLYRSNHWVDSRPEQSPQTPPEPSG 360
 DB 321 ARGVAVPHGDHYHFIPIYSQMSLEERIIARIIPLYRSNHWVDSRPEQSPQTPPEPSG 380
 QY 361 POPAPNLKIDNSLSVQLVRKVGEGYVPEEKIGISRYVFAKDLPSKLVESKLSKQES 420
 DB 381 POPAPNLKIDNSLSVQLVRKVGEGYVPEEKIGISRYVFAKDLPSKLVESKLSKQES 440

QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYLKDEDIVNEVKGKGVYIKVDGKYVYVLKDA 120
 DB 61 SHGDHYHYNGKVPYDAIISELLMKDPNYLKDEDIVNEVKGKGVYIKVDGKYVYVLKDA 120
 QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
 DB 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
 QY 181 GDYIVPHGDHYHYIPKNELASASLAAAEAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
 DB 181 GDYIVPHGDHYHYIPKNELASASLAAAEAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
 QY 241 VSNPGTTNTNTSNNTNSQASQNDSDLLKQLYKPLSORHVESDGLVDPQAITSRT 300
 DB 241 VSNPGTTNTNTSNNTNSQASQNDSDLLKQLYKPLSORHVESDGLVDPQAITSRT 300
 QY 301 ARGVAVPHGDHYHYIPYSQMSSELSERIARIIPLYRSNHNWVPSRQPSQPPTPEPSPG 360
 DB 301 ARGVAVPHGDHYHYIPYSQMSSELSERIARIIPLYRSNHNWVPSRQPSQPPTPEPSPG 360
 QY 361 POPAPNLKIDSNLSLVQVRKVGEGYVEEKGISRVYFAKDLPSVTKMLESKLSQES 420
 DB 361 POPAPNLKIDSNLSLVQVRKVGEGYVEEKGISRVYFAKDLPSVTKMLESKLSQES 420
 QY 421 VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLRLNDESTN 480
 DB 421 VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLRLNDESTN 480
 QY 481 KEKLVDDLLAFAPITPBERLKGKNSQIETVEDVRIAQLADKYTTSDGYIFDEHDIISD 540
 DB 481 KEKLVDDLLAFAPITPBERLKGKNSQIETVEDVRIAQLADKYTTSDGYIFDEHDIISD 540
 QY 541 EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPSPDADVKANPTGDSAAIY 600
 DB 541 EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPSPDADVKANPTGDSAAIY 600
 QY 601 NRKVGKRIPLVRLPYMVEHTVEYKNGNLIIPKHDHYHNIKFAWFDHTTYKAPNGYTTLED 660
 DB 601 NRKVGKRIPLVRLPYMVEHTVEYKNGNLIIPKHDHYHNIKFAWFDHTTYKAPNGYTTLED 660
 QY 661 LPATIKYVVEHDPDRPHNSDGNWGNASHVLGKKDHSDDPNKFNKADBEPEVETPAEPEVP 720
 DB 661 LPATIKYVVEHDPDRPHNSDGNWGNASHVLGKKDHSDDPNKFNKADBEPEVETPAEPEVP 720
 QY 721 QVTEKVEAQLKEAEVLLAKVTDSLLKANATETTLAQRNLTQIMDNNSIMAEAEKLLA 780
 DB 721 QVTEKVEAQLKEAEVLLAKVTDSLLKANATETTLAQRNLTQIMDNNSIMAEAEKLLA 780
 QY 781 LLKGSNPSSVSKPKIN 796
 DB 781 LLKGSNPSSVSKPKIN 796

RESULT 4
 ID ABU01598
 XX ABU01598 standard; protein; 802 AA.
 XX
 XX ABU01598;
 DT 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 XX S. pneumoniae type 4 strain protein from coding region #1174.
 DE
 DE
 KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae; type 4 strain.
 XX
 PN WQ200277021-A2.
 XX
 PD 03-OCT-2002.

XX PF 27-MAR-2002; 2002WO-IB002163.
 XX PR 27-MAR-2001; 2001GB-00007658.
 XX PA (CHIR-) CHIRON SPA.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Masignani V, Tettelin H, Fraser C;
 XX DR WPI; 2003-040579/03.
 XX DR N-PSDB; ABX06886.
 XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX PS Claim 1; SEQ ID NO 2348; 56pp; English.
 XX
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB556454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence,
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX SQ Sequence 802 AA;

Query Match 87.3%; Score 695; DB 6; Length 802;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
 DB 7 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 66
 QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYLKDEDIVNEVKGKGVYIKVDGKYVYVLKDA 120
 DB 67 SHGDHYHYNGKVPYDAIISELLMKDPNYLKDEDIVNEVKGKGVYIKVDGKYVYVLKDA 126
 QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
 DB 127 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 186
 QY 181 GDYIVPHGDHYHYIPKNELASASLAAAEAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
 DB 187 GDYIVPHGDHYHYIPKNELASASLAAAEAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 246

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX WPI; 2002-479261/51.
 DR N-PSDB; ABQ84819.
 XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 XX infection.
 XX Claim 11; Page 27; 70pp; English.
 XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54569. The S.
 CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcus infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention
 XX
 XX Sequence 796 AA;
 SQ
 Query Match 99.9%; Score 795; DB 5; Length 796;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
 Db 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
 QY 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDIEDIVNEVKGGYIVKDGKYYVYLKDA 120
 Db 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDIEDIVNEVKGGYIVKDGKYYVYLKDA 120
 QY 121 AHADNVRTKEENRQKQESQHQREGTTPRNDGAVALARSGRYTTDDGVIFNASDIIEDT 180
 Db 121 AHADNVRTKEENRQKQESQHQREGTTPRNDGAVALARSGRYTTDDGVIFNASDIIEDT 180
 QY 181 GDAYIVPHGDHYHYIPKNELSASELAFAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 240
 Db 181 GDAYIVPHGDHYHYIPKNELSASELAFAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 240
 QY 241 VSNPGTNTNTSNNSTNSQASQSDNIDSLKQLYKPLSQRHVESDGLVFDPAQITST 300
 Db 241 VSNPGTNTNTSNNSTNSQASQSDNIDSLKQLYKPLSQRHVESDGLVFDPAQITST 300
 QY 301 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
 Db 301 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
 QY 361 POPAPNLKIDSNSLSVQLVRKVGEGVPEEKISRYVFAKDLPSRTVKNLESKLSKQES 420
 Db 361 POPAPNLKIDSNSLSVQLVRKVGEGVPEEKISRYVFAKDLPSRTVKNLESKLSKQES 420
 QY 421 VSHLTAKENVAAPROEFYDKAYNLLTEAHKALFYKNGRNSDFQALDKLLERLNDESTN 480
 Db 421 VSHLTAKENVAAPROEFYDKAYNLLTEAHKALFYKNGRNSDFQALDKLLERLNDESTN 480
 QY 481 KEKLVDDLLAFLAPIHPERLGNPSQIEYTEDVRAIQAADKYTTSDGYIFDEHDIISD 540
 Db 481 KEKLVDDLLAFLAPIHPERLGNPSQIEYTEDVRAIQAADKYTTSDGYIFDEHDIISD 540
 QY 541 EGDAYVTPHMGSHWIGKSLSDKEKVAQAAYTKEGILPPSPDADVKANPTGDSAAIY 600
 Db 541 EGDAYVTPHMGSHWIGKSLSDKEKVAQAAYTKEGILPPSPDADVKANPTGDSAAIY 600
 QY 601 NRVKGEKRIPLVRLPYMVEHTVEKGNLIIIPKHQVHNITKEAWEDDHTYKAPNGYITLED 660
 Db 601 NRVKGEKRIPLVRLPYMVEHTVEKGNLIIIPKHQVHNITKEAWEDDHTYKAPNGYITLED 660

QY 661 LFIATIKYVVEHDERPHSNDGWNASEHVLGKKHSDPNKNFKADEEVEETPAEPVFP 720
 Db 661 LFIATIKYVVEHDERPHSNDGWNASEHVLGKKHSDPNKNFKADEEVEETPAEPVFP 720
 QY 721 QVETEKVRAQLKEAEVLLAKVTDSSI KANATETLAGLRNNLTLOIMDNNSINAEAEKLLA 780
 Db 721 QVETEKVRAQLKEAEVLLAKVTDSSI KANATETLAGLRNNLTLOIMDNNSINAEAEKLLA 780
 QY 781 LLKGSNPSSVSKEKIN 796
 Db 781 LLKGSNPSSVSKEKIN 796
 RESULT 3
 ADC45137
 ID ADC45137 standard; protein; 796 AA.
 AC ADC45137;
 DT 18-DEC-2003 (first entry)
 XX S. pneumoniae antigenic protein SP036.
 XX Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
 XX Streptococcus pneumoniae.
 XX US6573082-B1.
 XX 03-JUN-2003.
 XX 28-MAR-2000; 2000US-00536784.
 PR 31-OCT-1996; 96US-0029960P.
 PR 30-OCT-1997; 97US-00961083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX WPI; 2003-764574/72.
 DR N-PSDB; ADC45136.
 XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
 PT useful for producing vaccines for prevention or attenuation of infection
 PT by Streptococcus pneumoniae.
 XX Example 1; SEQ ID NO 56; 58pp; English.
 XX The invention relates to an isolated polynucleotide consisting of a
 CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
 CC antigens. Also included are making a recombinant vector by inserting the
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
 CC antigen nucleic acids are useful as probes for use in diagnostic methods
 CC for detecting S. pneumoniae gene expression. The present sequence
 CC represents an S. pneumoniae antigenic protein.
 XX
 XX Sequence 796 AA;
 SQ
 Query Match 99.9%; Score 795; DB 7; Length 796;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
 Db 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60

```

XX AC AAW55090;
XX DT 02-OCT-1998 (first entry)
XX DE Streptococcus pneumoniae SP0036 protein.
XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX KW detection; pneumonia; otitis media; meningitis.
XX OS Streptococcus pneumoniae.
XX FH Key Location/Qualifiers
XX FT Misc-difference 456
XX FT /label= unknown
XX FT /note= "encoded by GNA"
XX FN W09818930-A2.
XX XX 07-MAY-1998.
XX PD 30-OCT-1997; 97WO-US019422.
XX PF 31-OCT-1996; 96US-0029960P.
XX PR (HUMA-) HUMAN GENOME SCI INC.
XX PA Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX PI WPI; 1998-272224/24.
XX DR N-PSDB; AAV27351.
XX XX
XX PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
XX PT - or their epitope-containing fragments, useful in protective or
XX PT therapeutic vaccines, and for diagnosis.
XX XX
XX PS Claim 11; Page 59-60; 118pp; English.
XX XX
XX CC The present sequence represents a protein from Streptococcus pneumoniae.
XX CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX CC can be useful in vaccines for inducing protective antibodies against
XX CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX CC are used to detect Streptococcus infection (by usual hybridisation or
XX CC amplification methods), also for isolating Streptococcus genes or their
XX CC allelic variants. The protein can be used similarly to detect specific
XX CC antibodies in standard immunoassays, especially for diagnosing or
XX CC monitoring infections. Antibodies which bind the protein are used to
XX CC detect corresponding antigens, to purify the protein and for passive
XX CC immunisation (optionally coupled to a toxin). Vaccines are administered,
XX CC e.g. by injection, orally or through the skin, typically at 0.01-1000
XX CC (especially 10-300) mu g/ml per dose
XX XX
XX SQ Sequence 796 AA;
XX
Query Match 99.9%; Score 795; DB 2; Length 796;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYELGLYQARTVKNRVSVIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
DB 1 SYELGLYQARTVKNRVSVIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
QY 61 SHGDHYHYNGKYPYDAIISEELLMDPNYKLDIEDIVNEVKGGYIVKDGKYYVYLKDA 120
DB 61 SHGDHYHYNGKYPYDAIISEELLMDPNYKLDIEDIVNEVKGGYIVKDGKYYVYLKDA 120
QY 121 AHADNVRTKEEINRQKQESHQREGTTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
DB 121 AHADNVRTKEEINRQKQESHQREGTTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
QY 181 GDAYIVPHGDHYHYIPKNELSASELAEEAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240

```

```

Db 181 GDAYIVPHGDHYHYIPKNELSASELAEEAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
QY 241 VSNPCTTNTNTSNNSTNSQASQNDIDSLKQYKLPLOSORHVESDGLVFDPAQITSR 300
Db 241 VSNPCTTNTNTSNNSTNSQASQNDIDSLKQYKLPLOSORHVESDGLVFDPAQITSR 300
QY 301 ARGVAVPHGDHYHYIPYSQMSLEERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
Db 301 ARGVAVPHGDHYHYIPYSQMSLEERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
QY 361 POPAPNLKIDNSLSLVQVVKVGEVYVREKIGISRVYFAKDLPSFTVKLESKLSQES 420
Db 361 POPAPNLKIDNSLSLVQVVKVGEVYVREKIGISRVYFAKDLPSFTVKLESKLSQES 420
QY 421 VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db 421 VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
QY 481 KEKLVDDLLAFAPITHPERLKGPNISQIETEDVRIAQLADKYTTSDDGYIFDEHDIISD 540
Db 481 KEKLVDDLLAFAPITHPERLKGPNISQIETEDVRIAQLADKYTTSDDGYIFDEHDIISD 540
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKGILPPSPDADVKANPTGDSAAIY 600
Db 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKGILPPSPDADVKANPTGDSAAIY 600
QY 601 NRKGEKRIPLVRLPYMVEHTVEVQGNLIIPHKDHVHNIKFAWFDHTYKAPNGYTTLED 660
Db 601 NRKGEKRIPLVRLPYMVEHTVEVQGNLIIPHKDHVHNIKFAWFDHTYKAPNGYTTLED 660
QY 661 LFATIKYVVEHPDPERPHSNDGNGNASEHVLGKDHSEDPNKNPKADEEPVEETPAEPEVP 720
Db 661 LFATIKYVVEHPDPERPHSNDGNGNASEHVLGKDHSEDPNKNPKADEEPVEETPAEPEVP 720
QY 721 QVETEKVEAQLKEAEVLLAKVTDSLKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
Db 721 QVETEKVEAQLKEAEVLLAKVTDSLKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
QY 781 LLKGSNPSSVSKEKIN 796
Db 781 LLKGSNPSSVSKEKIN 796
RESULT 2
ABP54584
ID ABP54584 standard; protein; 796 AA.
XX ABP54584;
XX AC ABP54584;
XX DT 04-SEP-2002 (first entry)
XX DE S. pneumoniae SP036 protein sequence SEQ ID NO:56.
XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX KW antibacterial; Streptococcal infection; detection.
XX OS Streptococcus pneumoniae.
XX PN US2002061545-A1.
XX PD 23-MAY-2002.
XX PF 22-JAN-2001; 2001US-00765272.
XX PR 30-OCT-1997; 97US-00961083.
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.

```

| Result No. | Score | Query | | DB | ID | Description |
|------------|-------|-------|--------|----|-----------|---------------------|
| | | Match | Length | | | |
| 1 | 795 | 99.9 | 796 | 2 | AAW55090 | Aaw55090 Streptococ |
| 2 | 795 | 99.9 | 796 | 5 | ABP54584 | Abp54584 S. pneumo |
| 3 | 795 | 99.9 | 796 | 7 | ADC45137 | Adc45137 S. pneumo |
| 4 | 695 | 87.3 | 802 | 6 | ABU01598 | Abu01598 S. pneumo |
| 5 | 695 | 87.3 | 819 | 3 | AAB01468 | Aab01468 Recombina |
| 6 | 80 | 10.1 | 334 | 3 | AAAB12728 | Aab12728 Streptococ |
| 7 | 80 | 10.1 | 334 | 5 | AAU84027 | Aau84027 Truncated |
| 8 | 80 | 10.1 | 334 | 7 | ABM18808 | Abm18808 S. pneumo |
| 9 | 80 | 10.1 | 690 | 3 | AAU12745 | Aau12745 Streptococ |
| 10 | 80 | 10.1 | 690 | 5 | AAU84093 | Aau84093 Truncated |
| 11 | 80 | 10.1 | 690 | 7 | ABM18826 | Abm18826 S. pneumo |
| 12 | 80 | 10.1 | 821 | 3 | AAAB12766 | Aab12766 Streptococ |
| 13 | 80 | 10.1 | 821 | 3 | AAAB12727 | Aab12727 Streptococ |
| 14 | 80 | 10.1 | 821 | 5 | AAU84026 | Aau84026 Truncated |
| 15 | 80 | 10.1 | 821 | 7 | ABM18807 | Abm18807 S. pneumo |
| 16 | 80 | 10.1 | 840 | 3 | AAAB12716 | Aab12716 Streptococ |
| 17 | 80 | 10.1 | 840 | 5 | AAU75933 | Aau75933 Streptococ |
| 18 | 80 | 10.1 | 840 | 7 | ABM18797 | Abm18797 S. pneumo |
| 19 | 60 | 7.5 | 805 | 3 | AAAB12764 | Aab12764 Streptococ |
| 20 | 60 | 7.5 | 807 | 3 | AAAB12765 | Aab12765 Streptococ |
| 21 | 60 | 7.5 | 811 | 3 | AAAB12761 | Aab12761 Streptococ |
| 22 | 60 | 7.5 | 811 | 3 | AAAB12762 | Aab12762 Streptococ |
| 23 | 60 | 7.5 | 811 | 3 | AAAB12760 | Aab12760 Streptococ |
| 24 | 60 | 7.5 | 811 | 3 | AAAB12763 | Aab12763 Streptococ |
| 25 | 60 | 7.5 | 816 | 3 | AAAB12758 | Aab12758 Streptococ |

US-10-387-783-1
; Sequence 1, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 3.0
; SEQ ID NO 1
; LENGTH: 36
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Forward primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-10-387-783-1

Query Match 1.1%; Score 27; DB 16; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.0041;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCTTACGAGTTGGAGCTGTATCAAGC 27
| | | | | | | | | | | | | | | | | | | | |
Db 10 TTCTTACGAGTTGGAGCTGTATCAAGC 36

Search completed: October 1, 2004, 07:13:14
Job time : 1914 secs

; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 3
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Reverse primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-10-412-850-3

Query Match 1.3%; Score 30; DB 16; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 CCTTCATCTGTAAGTAAGGAAAAATAAAC 2389
DB 40 CCTTCATCTGTAAGTAAGGAAAAATAAAC 11

RESULT 22
US-10-387-783-3/c
; Sequence 3, Application US/10387783
; Publication No. US2004000531A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 3
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Reverse primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-10-387-783-3

Query Match 1.3%; Score 30; DB 16; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 CCTTCATCTGTAAGTAAGGAAAAATAAAC 2389
DB 40 CCTTCATCTGTAAGTAAGGAAAAATAAAC 11

RESULT 23
US-10-412-862-1

; Sequence 1, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1
; LENGTH: 36
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Forward primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-10-412-862-1

Query Match 1.1%; Score 27; DB 13; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.0041;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGC 27
DB 10 TTCTTACGAGTTGGGACTGTATCAAGC 36

RESULT 24
US-10-412-850-1
; Sequence 1, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1
; LENGTH: 36
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Forward primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-10-412-850-1

Query Match 1.1%; Score 27; DB 16; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.0041;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGC 27
DB 10 TTCTTACGAGTTGGGACTGTATCAAGC 36

RESULT 25

; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-5

Query Match 2.2%; Score 53; DB 16; Length 2531;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCACTGGGTACCAAGATTCAAG 1035
|||||
DB 1003 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCACTGGGTACCAAGATTCAAG 1055
|||||

RESULT 14

US-10-412-850-11
; Sequence 11, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-11

Query Match 2.2%; Score 53; DB 16; Length 2531;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCACTGGGTACCAAGATTCAAG 1035
|||||
DB 1000 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCACTGGGTACCAAGATTCAAG 1052
|||||

RESULT 15

US-10-387-783-5
; Sequence 5, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie E.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-5

Query Match 2.2%; Score 53; DB 16; Length 2531;

Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCACTGGGTACCAAGATTCAAG 1035
|||||
DB 1003 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCACTGGGTACCAAGATTCAAG 1055
|||||

RESULT 16

US-10-387-783-11
; Sequence 11, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie E.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-11

Query Match 2.2%; Score 53; DB 16; Length 2531;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCACTGGGTACCAAGATTCAAG 1035
|||||
DB 1000 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCACTGGGTACCAAGATTCAAG 1052
|||||

RESULT 17

US-09-884-465A-5
; Sequence 5, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2639
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-5

Query Match 2.2%; Score 53; DB 10; Length 2639;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCACTGGGTACCAAGATTCAAG 1035
|||||


```
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 243:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 243:
US-10-158-844-243

Query Match      2.2%; Score 53; DB 13; Length 2359;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
Db 1879 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1931

RESULT 10
US-09-769-787-206
; Sequence 206, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 206
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-206

Query Match      2.2%; Score 53; DB 10; Length 2481;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
Db 1003 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1055

RESULT 11
US-10-412-862-5
; Sequence 5, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
```

```
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-5

Query Match      2.2%; Score 53; DB 13; Length 2531;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
Db 1003 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1055

RESULT 12
US-10-412-862-11
; Sequence 11, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-11

Query Match      2.2%; Score 53; DB 13; Length 2531;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
Db 1000 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1052

RESULT 13
US-10-412-850-5
; Sequence 5, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
```

; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 2523
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-884-465A-3

Query Match 6.2%; Score 148; DB 10; Length 2523;
 Best Local Similarity 100.0%; Pred. No. 3.3e-67;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 699 CAAGAACAACTGGGTACCTCTGTAAGCAATCCAGGAAGTCAAAATCTACTACACAAGCA 758
 Db 758 CAAGAACAACTGGGTACCTCTGTAAGCAATCCAGGAAGTCAAAATCTACTACACAAGCA 817
 QY 759 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAAGTAAATGATGATCTCTTTGAAC 818
 Db 818 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAAGTAAATGATGATCTCTTTGAAC 877
 QY 819 AGCTCTACAACTGCCTTTGAGTCAACG 846
 Db 878 AGCTCTACAACTGCCTTTGAGTCAACG 905

RESULT 7
 US-09-884-465A-4
 ; Sequence 4, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; PRIOR FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 4
 ; LENGTH: 2647
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-884-465A-4

Query Match 6.2%; Score 148; DB 10; Length 2647;
 Best Local Similarity 100.0%; Pred. No. 3.4e-67;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 699 CAAGAACAACTGGGTACCTCTGTAAGCAATCCAGGAAGTCAAAATCTACTACACAAGCA 758
 Db 802 CAAGAACAACTGGGTACCTCTGTAAGCAATCCAGGAAGTCAAAATCTACTACACAAGCA 861
 QY 759 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAAGTAAATGATGATCTCTTTGAAC 818
 Db 862 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAAGTAAATGATGATCTCTTTGAAC 921
 QY 819 AGCTCTACAACTGCCTTTGAGTCAACG 846
 Db 922 AGCTCTACAACTGCCTTTGAGTCAACG 949

RESULT 8
 US-09-765-272-65
 ; Sequence 65, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 65:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2290 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 65:
 US-09-765-272-65
 Query Match 2.2%; Score 53; DB 9; Length 2290;
 Best Local Similarity 100.0%; Pred. No. 1e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATGGTACCAGATTCAAG 1035
 Db 944 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATGGTACCAGATTCAAG 996

RESULT 9
 US-10-158-844-243
 ; Sequence 243, Application US/10158844
 ; Publication No. US20040029118A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kunsch et al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-R
 ; COMPUTER: Dell Latitude Pentium 3
 ; OPERATING SYSTEM: Windows 98
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/158,844
 ; FILING DATE: 03-Jun-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/961,527
 ; FILING DATE: 1997-10-30

QY 481 AGGACGCTATATCAAGATGATGGTTATATCTTTAAATGCTTCTGATATCATAGAGGATAC 540
Db 3533 AGGACGCTATATCAAGATGATGGTTATATCTTTAAATGCTTCTGATATCATAGAGGATAC 3532
QY 541 TGGTATGCTTATATCGTTTCTCATGGAGATCATTTACCATTTACATTTCTTAAGAATGAGTT 600
Db 3593 TGGTATGCTTATATCGTTTCTCATGGAGATCATTTACCATTTACATTTCTTAAGAATGAGTT 3652
QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAGAGAGCTTCTCTATCTGCTCGAGGAATCTGTCAA 660
Db 3653 ATCAGCTAGCGAGTTGGCTGCTGCAGAGAGCTTCTCTATCTGCTCGAGGAATCTGTCAA 3712
QY 661 TTCAGAACCTTATCCCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
Db 3713 TTCAGAACCTTATCCCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 3772
QY 721 TGTAAAGCAATCCAGAACTACAAATATCTAACAGAGCAACAAAGCAACACTAACAGTCA 780
Db 3773 TGTAAAGCAATCCAGAACTACAAATATCTAACAGAGCAACAAAGCAACACTAACAGTCA 3832
QY 781 AGCAAGTCMAAGTAAATGACATTTGATGCTCTCTGAAACAGCTCTACAACTGCCTTTGAG 840
Db 3833 AGCAAGTCMAAGTAAATGACATTTGATGCTCTCTGAAACAGCTCTACAACTGCCTTTGAG 3892
QY 841 TCAACGACATGTAGAATCTGATGCCCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC 900
Db 3893 TCAACGACATGTAGAATCTGATGCCCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC 3952
QY 901 AGCTAGAGGTGTCAGTGCACAGGAGATCATTTACCTTATCCCTTACTCTCAAA 960
Db 3953 AGCTAGAGGTGTCAGTGCACAGGAGATCATTTACCTTATCCCTTACTCTCAAA 4012
QY 961 GTCTGAATTTGGAAGAACGAATCGCTCGTATTATTTCCCTTCTGTTATCTGTTCAAACTTG 1020
Db 4013 GTCTGAATTTGGAAGAACGAATCGCTCGTATTATTTCCCTTCTGTTATCTGTTCAAACTTG 4072
QY 1021 GGTAACGATTTAAGGCGAGAAACCAAGTCACAAAGCAACCGGAACTTAGTCCAGG 1080
Db 4073 GGTAACGATTTAAGGCGAGAAACCAAGTCACAAAGCAACCGGAACTTAGTCCAGG 4132
QY 1081 CCCGCACTTCCACCAATCTTAAATAGACTCAAAATCTTCTTGTGTTAGTCAAGTGGT 1140
Db 4133 CCCGCACTTCCACCAATCTTAAATAGACTCAAAATCTTCTTGTGTTAGTCAAGTGGT 4192
QY 1141 ACAGAAATTTGGGAGAGATGATGTTTCGAAGAAAGGGGATCTCTCGTTATGCTTTTC 1200
Db 4193 ACAGAAATTTGGGAGAGATGATGTTTCGAAGAAAGGGGATCTCTCGTTATGCTTTTC 4252
QY 1201 GAAAGATTTACCATCTGAATCTGTTAAATCTTTGAAGCAAGTTATCAAAAACAGAGAG 1260
Db 4253 GAAAGATTTACCATCTGAATCTGTTAAATCTTTGAAGCAAGTTATCAAAAACAGAGAG 4312
QY 1261 TGTTTCACACATTTAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTCNAATTAAGGTCG 1380
Db 4373 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTCNAATTAAGGTCG 4432
QY 1381 TAATTCGATTTCCAGCCTTAGACAAATTTATAGAACCGTTGAATGATGAATCGACTAA 1440
Db 4433 TAATTCGATTTCCAGCCTTAGACAAATTTATAGAACCGTTGAATGATGAATCGACTAA 4492
QY 1441 TAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCAACCAATTTACCATCCAGCG 1500
Db 4493 TAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCAACCAATTTACCATCCAGCG 4552
QY 1501 ACTTGGCAACCAAAATTTCTCAAAATGAGTATCTGAAGACGAAGTTCTGTTATCTCAAT 1560
Db 4553 ACTTGGCAACCAAAATTTCTCAAAATGAGTATCTGAAGACGAAGTTCTGTTATCTCAAT 4612
QY 1561 AGCTGATAAGTATACAAAGTCAGATGGTTACATTTTGTGTAACATGATATATCATGTA 1620

Db 4613 AGCTGATAAGTATACAAAGTCAGATGGTTACATTTTGTGTAACATGATATATCATGTA 4672
QY 1621 TGAAGGAGATGATATGATTAAGCGCTCATATGGGCCATAGTCACTTGGATTTGAAAAGTAG 1680
Db 4673 TGAAGGAGATGATATGATTAAGCGCTCATATGGGCCATAGTCACTTGGATTTGAAAAGTAG 4732
QY 1681 CTTTCTGATAAGGAAAAGTTGAGCTCAAGCTTATCTAAAGAAAAGGTATCTTACC 1740
Db 4733 CTTTCTGATAAGGAAAAGTTGAGCTCAAGCTTATCTAAAGAAAAGGTATCTTACC 4792
QY 1741 TCCATCTCCAGACGACATGTTAAAGCAAAATCCAACCTGGAGATAGTCCAGCAGCTTTTA 1800
Db 4793 TCCATCTCCAGACGACATGTTAAAGCAAAATCCAACCTGGAGATAGTCCAGCAGCTTTTA 4852
QY 1801 CAATCGTGTGAAAAGGGGAAAACGAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1860
Db 4853 CAATCGTGTGAAAAGGGGAAAACGAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 4912
QY 1861 TACAGTTGAGGTTTAAAAAACGTTAATTTGATTTTCTCATAGGATCATTTACCATATAT 1920
Db 4913 TACAGTTGAGGTTTAAAAAACGTTAATTTGATTTTCTCATAGGATCATTTACCATATAT 4972
QY 1921 TAAATTTGCTTGGTTTGTATGATCACATACAAAGCTCCAAATGGCTTATACCTTGAAGA 1980
Db 4973 TAAATTTGCTTGGTTTGTATGATCACATACAAAGCTCCAAATGGCTTATACCTTGAAGA 5032
QY 1981 TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCTTGACGAAACGCTCCACATTTATGA 2040
Db 5033 TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCTTGACGAAACGCTCCACATTTATGA 5092
QY 2041 TGGATGGGCAATGCGAGTACGATGTTAGGCAAGAAACACCTGCTGAGCCAGAGTCC 2100
Db 5093 TGGATGGGCAATGCGAGTACGATGTTAGGCAAGAAAGACCCCTGACGAACTCCATTAATGA 5152
QY 2101 TAAAGACTTCAAAAGCGGATGAAGAGCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
Db 5153 TAAAGACTTCAAAAGCGGATGAAGAGCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 5212
QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGCAAGTTTGTCTTGGAA 2220
Db 5213 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGCAAGTTTGTCTTGGAA 5272
QY 2221 AGTAACGGATTTCTAGTCTGAAAGCCATGCAAGAACTCTAGTCTGTTTACGAATAA 2280
Db 5273 AGTAACGGATTTCTAGTCTGAAAGCCATGCAAGAACTCTAGTCTGTTTACGAATAA 5332
QY 2281 TTTGACTCTTCAAAATTTATGATTAACAATAGTATCATGGCAGAGAGCAAGAAATTTACTTGC 2340
Db 5333 TTTGACTCTTCAAAATTTATGATTAACAATAGTATCATGGCAGAGAGCAAGAAATTTACTTGC 5392
QY 2341 GTTGTAAAGGAGTAATCTTCTATCTGTAAGTAAAGGAAAAATAAAC 2389
Db 5393 GTTGTAAAGGAGTAATCTTCTATCTGTAAGTAAAGGAAAAATAAAC 5441

RESULT 6

US-09-884-465A-3
; Sequence 3, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20

| | | | |
|----|------|---|------|
| Db | 1440 | TAATTCCTGATTTCCAAAGCCCTTAGACAAATTAATTAGAACCGCTTGAATGATGAATCGACTAA | 1499 |
| Qy | 1441 | TAAAGAAAAATTGGTAGATGATTAATTATGGCAATTCCTAGACCAATTAACCATCCAGAGCG | 1500 |
| Db | 1500 | TAAAGAAAAATTGGTAGATGATTAATTATGGCAATTCCTAGACCAATTAACCATCCAGAGCG | 1559 |
| Qy | 1501 | ACTTGGCAAAACCAAATTCCTCAAAATTGAGTATATCTGAAGACGAAGTTCGTATTGCTCAATTT | 1560 |
| Db | 1560 | ACTTGGCAAAACCAAATTCCTCAAAATTGAGTATATCTGAAGACGAAGTTCGTATTGCTCAATTT | 1619 |
| Qy | 1561 | AGCTGATAAGTATACAACGTCAGATGGTTACATTTTGTGAAACATGATATAATCAGTGA | 1620 |
| Db | 1620 | AGCTGATAAGTATACAACGTCAGATGGTTACATTTTGTGAAACATGATATAATCAGTGA | 1679 |
| Qy | 1621 | TGAAGGAGATGCATATATAACGCTCATATGGGCCATAGTCACTGGATTTGGAAAAAGATAG | 1680 |
| Db | 1680 | TGAAGGAGATGCATATATAACGCTCATATGGGCCATAGTCACTGGATTTGGAAAAAGATAG | 1739 |
| Qy | 1681 | CCTTTCTGATAAGAAAAAGTTGCAGCTCAAGCCTATACCTAAGAAAAAGGTATCCTACC | 1740 |
| Db | 1740 | CCTTTCTGATAAGAAAAAGTTGCAGCTCAAGCCTATACCTAAGAAAAAGGTATCCTACC | 1799 |
| Qy | 1741 | TCCATCTCCAGACCGCAGATGTTAAAGCAAAATCCAACCTGGAGATAGTGCACGACTATTTA | 1800 |
| Db | 1800 | TCCATCTCCAGACCGCAGATGTTAAAGCAAAATCCAACCTGGAGATAGTGCACGACTATTTA | 1859 |
| Qy | 1801 | CAATCGTGTGAAGGGGAAAAAGAAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA | 1860 |
| Db | 1860 | CAATCGTGTGAAGGGGAAAAAGAAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA | 1919 |
| Qy | 1861 | TACAGTTTGAGGTAAAAAACGGTAATTTGATTTATTCCTCATAAGGATCAATTAACCATTAAT | 1920 |
| Db | 1920 | TACAGTTTGAGGTAAAAAACGGTAATTTGATTTATTCCTCATAAGGATCAATTAACCATTAAT | 1979 |
| Qy | 1921 | TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGGAAGA | 1980 |
| Db | 1980 | TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGGAAGA | 2039 |
| Qy | 1981 | TTTTGTTGCGACGATTAAGTACTACGTAGAACACCTCGACGAACGTCACCATTTCTAATGA | 2040 |
| Db | 2040 | TTTTGTTGCGACGATTAAGTACTACGTAGAACACCTCGACGAACGTCACCATTTCTAATGA | 2099 |
| Qy | 2041 | TGGATGGGGCAATGCCAGTGAAGCATGTGTTAGGCAAGAAAAACACACAGTGAAGATCCAAA | 2100 |
| Db | 2100 | TGGATGGGGCAATGCCAGTGAAGCATGTGTTAGGCAAGAAAAACACACAGTGAAGATCCAAA | 2159 |
| Qy | 2101 | TAGAACTTTCAAGGGGATGAAGAGCCAGTAGAGAAACACCTCTGAGCCAGAGTCCC | 2160 |
| Db | 2160 | TAGAACTTTCAAGGGGATGAAGAGCCAGTAGAGAAACACCTCTGAGCCAGAGTCCC | 2219 |
| Qy | 2161 | TCAAGTTAGAGACTGAAAAGGTAGAAAGCCCAACTCAAGAAAGCAAGTTTTGCTTGGGAA | 2220 |
| Db | 2220 | TCAAGTTAGAGACTGAAAAGGTAGAAAGCCCAACTCAAGAAAGCAAGTTTTGCTTGGGAA | 2279 |
| Qy | 2221 | AGTAAACGATTTAGTCTGAAAGCCAATGCAACGAAACTCTAGCTGGTTTACGAATATA | 2280 |
| Db | 2280 | AGTAAACGATTTAGTCTGAAAGCCAATGCAACGAAACTCTAGCTGGTTTACGAATATA | 2339 |
| Qy | 2281 | TTTGTACTCTTCAAAATTATGGATACAATAGTATCATGCGAAGCAGAAAAATTTACTTGC | 2340 |
| Db | 2340 | TTTGTACTCTTCAAAATTATGGATACAATAGTATCATGCGAAGCAGAAAAATTTACTTGC | 2399 |
| Qy | 2341 | GTGTGTAAGAGGAGTAACTCTTCACTGTAGTAAGGAAAAATAAAC | 2389 |
| Db | 2400 | GTGTGTAAGAGGAGTAACTCTTCACTGTAGTAAGGAAAAATAAAC | 2448 |

APPLICANT: Kunsch et al.
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-R
COMPUTER: Dell Latitude Pentium 3
OPERATING SYSTEM: Windows 98
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/158,844
FILING DATE: 03-Jun-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/961,527
FILING DATE: 1997-10-30
APPLICATION NUMBER: US 60/029,960
FILING DATE: 1996-10-31
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Mark J.
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB340P1D1
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 8195 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 94:
US-10-158-844-94

| Query Match | 97.9%; | Score 2338; | DB 13; | Length 8195; |
|-----------------------|-----------------|---|-----------|--------------|
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 2388; | Conservative 0; | Mismatches 1; | Indels 0; | Gaps 0; |
| QY | 1 | TTCTTACGAGTGTGGGACTGTATCAAGCTAGAACGTTAAGGAAAAAATCGTGTTCCTA | 60 | |
| DB | 3053 | TTCTTACGAGTGTGGGACTGTATCAAGCTAGAACGTTAAGGAAAAAATCGTGTTCCTA | 3112 | |
| QY | 61 | TATAGATGGAACAACAGCGACGCAAAAAACGGAGAAATTGACTTCCTGATGAGGTTAGCAA | 120 | |
| DB | 3113 | TATAGATGGAACAACAGCGACGCAAAAAACGGAGAAATTGACTTCCTGATGAGGTTAGCAA | 3172 | |
| QY | 121 | GCGTGAGGAATCAATGCTCGACGAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC | 180 | |
| DB | 3173 | GCGTGAGGAATCAATGCTCGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC | 3232 | |
| QY | 181 | TTCACATGGCGACCACTATCATTTACAATGTAAGGTTCCATTATGACGCTATCATCAG | 240 | |
| DB | 3233 | TTCACATGGCGACCACTATCATTTACAATGTAAGGTTCCATTATGACGCTATCATCAG | 3292 | |
| QY | 241 | TGAAGAAATTACTCATGNAAGATCCAACTATAGCTTAAAGATGAGGATATTGTTAATGA | 300 | |
| DB | 3293 | TGAAGAAATTACTCATGAAGAATCCAACTATAGCTTAAAGATGAGGATATTGTTAATGA | 3352 | |
| QY | 301 | GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATCATATGTTTACCTTAAGGATGC | 360 | |
| DB | 3353 | GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATCATATGTTTACCTTAAGGATGC | 3412 | |
| QY | 361 | TGCCACGCGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG | 420 | |
| DB | 3413 | TGCCACGCGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG | 3472 | |
| QY | 421 | TCAACATCGTGAAGGTGGAACTCCAGAAAAACGATGGTGCCTGTGGCAGTTCGCA | 480 | |
| DB | 3473 | TCAACATCGTGAAGGTGGAACTCCAGAAACGATGGTGCCTGTGGCAGTTCGCA | 3532 | |

RESULT 5
US-10-158-844-94
; Sequence 94, Application US/10158844
; Publication NO. US20040029118A1
; GENERAL INFORMATION:

Db 2160 TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAAACACCTGCTGAGCCAGAGTCCC 2219
 QY 2161 TCAAGTAGAGACTGAAGAACTAGAGCCCAACTCAAGAGACGAGAGTTTGGTTCGAA 2220
 Db 2220 TCAAGTAGAGACTGAAGAACTAGAGCCCAACTCAAGAGACGAGAGTTTGGTTCGAA 2279
 QY 2221 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
 Db 2280 AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2339
 QY 2281 TTTGACTCTTCAAAATTATGGATAACAATAGTATCATGCGAGAGACGAGAAATTTACTTGC 2340
 Db 2340 TTTGACTCTTCAAAATTATGGATAACAATAGTATCATGCGAGAGACGAGAAATTTACTTGC 2399
 QY 2341 GTTCTTAAAGGAGTAACTCTTCACTCTGTAAGTAAGGAAATAAATTAAC 2389
 Db 2400 GTTCTTAAAGGAGTAACTCTTCACTCTGTAAGTAAGGAAATAAATTAAC 2448

RESULT 4
 US-10-387-783-9
 ; Sequence 9, Application US/10387783
 ; Publication No. US20040005331A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-683
 ; CURRENT APPLICATION NUMBER: US/10/387,783
 ; PRIOR FILING DATE: 2003-03-13
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 9
 ; LENGTH: 2451
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)..(2451)
 ; OTHER INFORMATION: n = a, c, t or g
 US-10-387-783-9

Query Match 100.0%; Score 2389; DB 16; Length 2451;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTTCTAGAGTGGGACTGTATCAAGCTAGACGAGTAAAGGAAATTAATCGTTTCCTA 60
 Db 60 TTTCTAGAGTGGGACTGTATCAAGCTAGACGAGTAAAGGAAATTAATCGTTTCCTA 119
 QY 61 TATAGATGGAACCAAGCGACGCAAAAACCGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
 Db 120 TATAGATGGAACCAAGCGACGCAAAAACCGAGAAATTTGACTCCTGATGAGGTTAGCAA 179
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGCAAGGCTATGTCAC 180
 Db 180 GCGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGCAAGGCTATGTCAC 239
 QY 181 TTCACATGGCGACCTATCATTTATACATGTAAGTTTCCTTATGAGCTATCATCAG 240
 Db 240 TTCACATGGCGACCTATCATTTATACATGTAAGTTTCCTTATGAGCTATCATCAG 299
 QY 241 TGAAGAAATTAATCATGAAGATCCAACTATAGCTTAAAGATGAGGATATGTTAATGA 300
 Db 300 TGAAGAAATTAATCATGAAGATCCAACTATAGCTTAAAGATGAGGATATGTTAATGA 359

QY 301 GGTCAAGGTTGGATATGTTTCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC 360
 Db 360 GGTCAAGGTTGGATATGTTTCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC 419
 QY 361 TGCCCAAGCGGATAACCTGCGTACAAAAGAGGAAATCAATCGACAAAACAAGAGCATAG 420
 Db 420 TGCCCAAGCGGATAACCTGCGTACAAAAGAGGAAATCAATCGACAAAACAAGAGCATAG 479
 QY 421 TCAACATCGTGAAGGTTGGAATCTCAAGAAACGATGGTGTGCTTGGCAGCTTCGCA 480
 Db 480 TCAACATCGTGAAGGTTGGAATCTCAAGAAACGATGGTGTGCTTGGCAGCTTCGCA 539
 QY 481 AGGACGCTATACAGATGATGTTTATATCTTAAATGCTTCTGATATCATAGAGGATAC 540
 Db 540 AGGACGCTATACAGATGATGTTTATATCTTAAATGCTTCTGATATCATAGAGGATAC 599
 QY 541 TGTGTGATGCTTATATGTTCTCATGAGATCATTAACCATTAATCTTAAGAAATGAGTT 600
 Db 600 TGTGTGATGCTTATATGTTCTCATGAGATCATTAACCATTAATCTTAAGAAATGAGTT 659
 QY 601 ATCAGTAGCGAGTGGTCTGCTGAGAGGCTTCTTATCTGGTCGAGAAATCTGTCAA 660
 Db 660 ATCAGTAGCGAGTGGTCTGCTGAGAGGCTTCTTATCTGGTCGAGAAATCTGTCAA 719
 QY 661 TTCAAGAACTATCGCGCACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
 Db 720 TTCAAGAACTATCGCGCACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 779
 QY 721 TGTAAAGCAATCCAGGAACCTACAAATACTAACAAAGCAACCAAGCAAACTAACAGTCA 780
 Db 780 TGTAAAGCAATCCAGGAACCTACAAATACTAACAAAGCAACCAAGCAAACTAACAGTCA 839
 QY 781 AGCAAGTCAAGTAATGACATGATAGTCTTCTGAAACAGCTCTCAAACTGCTTTGAG 840
 Db 840 AGCAAGTCAAGTAATGACATGATAGTCTTCTGAAACAGCTCTCAAACTGCTTTGAG 899
 QY 841 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGACAAATCACAGTCCGAC 900
 Db 900 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGACAAATCACAGTCCGAC 959
 QY 901 AGCTAGAGTGTTCAGTGCCACACGAGATCATTAACCACTTCTATCCCTTACTCTCAAA 960
 Db 960 AGCTAGAGTGTTCAGTGCCACACGAGATCATTAACCACTTCTATCCCTTACTCTCAAA 1019
 QY 961 GTCTGAATTTGGAAGACGATCGCTCGTATATTCCTTCTGTTATCGTTTCAACCATTCG 1020
 Db 1020 GTCTGAATTTGGAAGACGATCGCTCGTATATTCCTTCTGTTATCGTTTCAACCATTCG 1079
 QY 1021 GGTACACAGATTCGAAGGCGAGAACCAAGTCCACAAACGACTCCGGAACCTAGTCCAG 1080
 Db 1080 GGTACACAGATTCGAAGGCGAGAACCAAGTCCACAAACGACTCCGGAACCTAGTCCAG 1139
 QY 1081 CCGGAACTGCAACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT 1140
 Db 1140 CCGGAACTGCAACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT 1199
 QY 1141 ACGAAAAGTTGGGAGGATATGTAATCGAAGAAAGGCTCTCTCGTTATGCTTTTC 1200
 Db 1200 ACGAAAAGTTGGGAGGATATGTAATCGAAGAAAGGCTCTCTCGTTATGCTTTTC 1259
 QY 1201 GAAAGATTTTACCCTGTAACCTGTTTAAATCTTGAAGCAAGTTTATCAAAACAAGAGAG 1260
 Db 1260 GAAAGATTTTACCCTGTAACCTGTTTAAATCTTGAAGCAAGTTTATCAAAACAAGAGAG 1319
 QY 1261 TGTTCACACACTTTAATGCTTAAATAGAAAATGTTGCTCTCTGAGCAAGAAATTTTA 1320
 Db 1320 TGTTCACACACTTTAATGCTTAAATAGAAAATGTTGCTCTCTGAGCAAGAAATTTTA 1379
 QY 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATTAAGGTCG 1380
 Db 1380 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATTAAGGTCG 1439
 QY 1381 TAAATCTGATTTCCCAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGATCGACTAA 1440

| Query Match | | | | | | | | | |
|---|------|--|------|--|--|--|--|--|--|
| Best Local Similarity 100.0%; Score 2389; DB 16; Length 2451; | | | | | | | | | |
| Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | | | | | | | | |
| QY | 1 | TTCTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAAGGAAATAATCTGTGTTTCCTA | 60 | | | | | | |
| DB | 60 | TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAAGGAAATAATCTGTGTTTCCTA | 119 | | | | | | |
| QY | 61 | TATAGATGGAATAAACAAGCGACGCAAAAACCGAGAAATTTGATCTCTGTAGAGTTAGCAA | 120 | | | | | | |
| DB | 120 | TATAGATGGAATAAACAAGCGACGCAAAAACCGAGAAATTTGATCTCTGTAGAGTTAGCAA | 179 | | | | | | |
| QY | 121 | GGGTGAAGGANTCAATGCTGAGCAATCGTATCAAGATAACAGACCAAGGCTATGTAC | 180 | | | | | | |
| DB | 180 | GGGTGAAGGANTCAATGCTGAGCAATCGTATCAAGATAACAGACCAAGGCTATGTAC | 239 | | | | | | |
| QY | 181 | TTCAATCGGACCACTATCATTTATTAATGTTAAGGTTTCCTATGACGCTATCATCAG | 240 | | | | | | |
| DB | 240 | TTCAATCGGACCACTATCATTTATTAATGTTAAGGTTTCCTATGACGCTATCATCAG | 299 | | | | | | |
| QY | 241 | TGAAGATTTACTCATGAAAGATCCAAACTATTAAGCTAAAGATGAGGATATTTGTTAATGA | 300 | | | | | | |
| DB | 300 | TGAAGATTTACTCATGAAAGATCCAAACTATTAAGCTAAAGATGAGGATATTTGTTAATGA | 359 | | | | | | |
| QY | 301 | GGTGAAGGTTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC | 360 | | | | | | |
| DB | 360 | GGTGAAGGTTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC | 419 | | | | | | |
| QY | 361 | TGCCACGCGGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAACAGAGCATAG | 420 | | | | | | |
| DB | 420 | TGCCACGCGGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAACAGAGCATAG | 479 | | | | | | |
| QY | 421 | TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGGTGTGTTGCTTGGGACGTTTCGCA | 480 | | | | | | |
| DB | 480 | TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGGTGTGTTGCTTGGGACGTTTCGCA | 539 | | | | | | |
| QY | 481 | AGGACGCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC | 540 | | | | | | |
| DB | 540 | AGGACGCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC | 599 | | | | | | |
| QY | 541 | TGCTGATGCTTATATCTGTTCTCATGGAGATCATTTACCATTTACATTTCTTAAGAAATGAGTT | 600 | | | | | | |
| DB | 600 | TGCTGATGCTTATATCTGTTCTCATGGAGATCATTTACCATTTACATTTCTTAAGAAATGAGTT | 659 | | | | | | |
| QY | 601 | ATCAGCTACGAGTTGGCTGTGCAAGAGCCTTCTATCTGTCGAGGAAATCTGTCAA | 660 | | | | | | |
| DB | 660 | ATCAGCTACGAGTTGGCTGTGCAAGAGCCTTCTATCTGTCGAGGAAATCTGTCAA | 719 | | | | | | |
| QY | 661 | TTCAAGAACTTATCGCGCAAAATAGCGATACACTTCAAGAACAAACTGGGTACCTTC | 720 | | | | | | |
| DB | 720 | TTCAAGAACTTATCGCGCAAAATAGCGATACACTTCAAGAACAAACTGGGTACCTTC | 779 | | | | | | |
| QY | 721 | TGTAAAGCAATCCAGGAATCTACAAATCTAACCAAGCAACAGCAACCACTAACAGTCA | 780 | | | | | | |
| DB | 780 | TGTAAAGCAATCCAGGAATCTACAAATCTAACCAAGCAACAGCAACCACTAACAGTCA | 839 | | | | | | |
| QY | 781 | AGCAAGTAAAGTAATGATGATGATCTCTTTGAAACAGCTCTACAACTGCTTTGAG | 840 | | | | | | |
| DB | 840 | AGCAAGTAAAGTAATGATGATGATCTCTTTGAAACAGCTCTACAACTGCTTTGAG | 899 | | | | | | |
| QY | 841 | TCAACGACATGTAGATCTGATGGCTGTCTTTGATCCAGCAAAATCACAGTCCGAAC | 900 | | | | | | |
| DB | 900 | TCAACGACATGTAGATCTGATGGCTGTCTTTGATCCAGCAAAATCACAGTCCGAAC | 959 | | | | | | |
| QY | 901 | AGCTAGAGTGTGCGATGCGCACAGGAGATCATTTACCACTTCATCCCTTACTCTCAAA | 960 | | | | | | |
| DB | 960 | AGCTAGAGTGTGCGATGCGCACAGGAGATCATTTACCACTTCATCCCTTACTCTCAAA | 1019 | | | | | | |
| QY | 961 | GTCTGAATTTGGAAGAACGAATCGCTCGTATATTCCTTCGTTATCGTTCAAAACCATTC | 1020 | | | | | | |
| DB | 1020 | GTCTGAATTTGGAAGAACGAATCGCTCGTATATTCCTTCGTTATCGTTCAAAACCATTC | 1079 | | | | | | |

| QY | 1021 | GGTACGATTTCAAGGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG | 1080 |
|----|------|--|------|
| DB | 1080 | GGTACGATTTCAAGGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG | 1139 |
| QY | 1081 | CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT | 1140 |
| DB | 1140 | CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT | 1199 |
| QY | 1141 | ACGAAAGTTGGGGAAGGATATGTTTCGAAGAAAGGCACTCTCTCGTATGTCTTTGC | 1200 |
| DB | 1200 | ACGAAAGTTGGGGAAGGATATGTTTCGAAGAAAGGCACTCTCTCGTATGTCTTTGC | 1259 |
| QY | 1201 | GAAAGATTTACCATCTGAAACTGTTTAAATAATCTTGAAGCAAGTTATCAAAAAGAGAG | 1260 |
| DB | 1260 | GAAAGATTTACCATCTGAAACTGTTTAAATAATCTTGAAGCAAGTTATCAAAAAGAGAG | 1319 |
| QY | 1261 | TGTTTCAACACTTTAACTGCTTAAAAAAGAAATGTTGCTCTCTGACCAAGAAATTTTA | 1320 |
| DB | 1320 | TGTTTCAACACTTTAACTGCTTAAAAAAGAAATGTTGCTCTCTGACCAAGAAATTTTA | 1379 |
| QY | 1321 | TGATAAAGCATATAATCTGTTAACTGAGCTCATAAAGCCTTGTGTTGNAATAAAGGCTG | 1380 |
| DB | 1380 | TGATAAAGCATATAATCTGTTAACTGAGCTCATAAAGCCTTGTGTTGNAATAAAGGCTG | 1439 |
| QY | 1381 | TAAATCTGATTTCCAAAGCCTTAGACAAATATTAGAACGCTTGAATGATGAAATCGACTAA | 1440 |
| DB | 1440 | TAAATCTGATTTCCAAAGCCTTAGACAAATATTAGAACGCTTGAATGATGAAATCGACTAA | 1499 |
| QY | 1441 | TAAAGAAAATTTGGTAGATGATTTATGTCATTTCTAGCACCAATTAACCCATCCAGAGG | 1500 |
| DB | 1500 | TAAAGAAAATTTGGTAGATGATTTATGTCATTTCTAGCACCAATTAACCCATCCAGAGG | 1559 |
| QY | 1501 | ACTTCGCAAAACCAAAATCTTCAAAATGAGTATCTGAAGACGAAAGTTTCGTTATGCTCAAT | 1560 |
| DB | 1560 | ACTTCGCAAAACCAAAATCTTCAAAATGAGTATCTGAAGACGAAAGTTTCGTTATGCTCAAT | 1619 |
| QY | 1561 | AGCTGATAGTATACAAAGCTCAGATGGTTTACATTTTGTGATGAACATGATATATAGTGA | 1620 |
| DB | 1620 | AGCTGATAGTATACAAAGCTCAGATGGTTTACATTTTGTGATGAACATGATATATAGTGA | 1679 |
| QY | 1621 | TGAAGAGATGATATGTHAACGCTCATATGGCCATAGTCTGATGATTTGGAAAGATAG | 1680 |
| DB | 1680 | TGAAGAGATGATATGTHAACGCTCATATGGCCATAGTCTGATGATTTGGAAAGATAG | 1739 |
| QY | 1681 | CTTTCTGTATAGGAAAAAGTTGCGAGCTCAAGCCTATATACTAAAGAAAAGTATCTTACC | 1740 |
| DB | 1740 | CTTTCTGTATAGGAAAAAGTTGCGAGCTCAAGCCTATATACTAAAGAAAAGTATCTTACC | 1799 |
| QY | 1741 | TCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA | 1800 |
| DB | 1800 | TCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA | 1859 |
| QY | 1801 | CAATCGTGTGAAGGGAAGAAAGGATTCCTCGTTCCGACTTCCATATATGTTCCAGCA | 1860 |
| DB | 1860 | CAATCGTGTGAAGGGAAGAAAGGATTCCTCGTTCCGACTTCCATATATGTTCCAGCA | 1919 |
| QY | 1861 | TACAGTTGAGTTTAAAAACGGTAAATTTGATTTTCTCTATAAGGATCATTTACCATAAT | 1920 |
| DB | 1920 | TACAGTTGAGTTTAAAAACGGTAAATTTGATTTTCTCTATAAGGATCATTTACCATAAT | 1979 |
| QY | 1921 | TAAATTTGCTGTTTGTGATGATCACAATAAAGCTCCAAATGGCTATACCTTGAAGA | 1980 |
| DB | 1980 | TAAATTTGCTGTTTGTGATGATCACAATAAAGCTCCAAATGGCTATACCTTGAAGA | 2039 |
| QY | 1981 | TTTGTGTTGCGAGATTAAGTACTACCTAGAACACCTGAGCAAGCTCCCATTTCTAATGA | 2040 |
| DB | 2040 | TTTGTGTTGCGAGATTAAGTACTACCTAGAACACCTGAGCAAGCTCCCATTTCTAATGA | 2099 |
| QY | 2041 | TGATGGGGCAATGCCAGTGAGCATGTGTTTAGCAAGAAAGACCAAGTGAAGATCCAAA | 2100 |
| DB | 2100 | TGATGGGGCAATGCCAGTGAGCATGTGTTTAGCAAGAAAGACCAAGTGAAGATCCAAA | 2159 |
| QY | 2101 | TAAAGAACTTCAAGCGGATGAAGGCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC | 2160 |

QY 661 TTCAAGAACCTATCGCGACAAAATAGCGATAACACACTTCAAGAACAAACTGGGTACCTTC 720
 Db 720 TTCAAGAACCTATCGCGACAAAATAGCGATAACACACTTCAAGAACAAACTGGGTACCTTC 779
 QY 721 TGTAAAGCAATCCAGGAATCAAAATCTAACTAACAAGCAACCAACAGCAACACTAACAGTCA 780
 Db 780 TGTAAAGCAATCCAGGAATCAAAATCTAACTAACAAGCAACCAACAGCAACACTAACAGTCA 839
 QY 781 AGCAAGTCAAGTAATGACATGATGATGCTCTTGAACAGCTCTACAAAGTCCCTTTGAG 840
 Db 840 AGCAAGTCAAGTAATGACATGATGATGCTCTTGAACAGCTCTACAAAGTCCCTTTGAG 899
 QY 841 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCACCAAAATCACAAGTCCGAAC 900
 Db 900 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCACCAAAATCACAAGTCCGAAC 959
 QY 901 AGCTAGAGGTGTTGACGTGCGACACAGGAGATCAATACCACTTCATCCCTTACTCTCAAT 960
 Db 960 AGCTAGAGGTGTTGACGTGCGACACAGGAGATCAATACCACTTCATCCCTTACTCTCAAT 1019
 QY 961 GTCTGAATTTGGAAGCAAGTCCGCTCGTATTAATTCCTGTTTATCGTTCAACCAATG 1020
 Db 1020 GTCTGAATTTGGAAGCAAGTCCGCTCGTATTAATTCCTGTTTATCGTTCAACCAATG 1079
 QY 1021 GGTACCAAGATTCAAGGCGCAGAAACAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080
 Db 1080 GGTACCAAGATTCAAGGCGCAGAAACAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1139
 QY 1081 CCGCAACCTGCGACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGGT 1140
 Db 1140 CCGCAACCTGCGACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGGT 1199
 QY 1141 ACAGAAAGTTGGGGAAGGATGATGATTCGAAGAAAGGCTCTCTGTTTATGCTTTGTC 1200
 Db 1200 ACAGAAAGTTGGGGAAGGATGATGATTCGAAGAAAGGCTCTCTGTTTATGCTTTGTC 1259
 QY 1201 GAAAGATTACCACTGTAACCTGTAACCACTGTAACCACTGTAACCACTGTAACCACTGTA 1260
 Db 1260 GAAAGATTACCACTGTAACCTGTAACCACTGTAACCACTGTAACCACTGTAACCACTGTA 1319
 QY 1261 TGTTTCACACATTTAACTGCTGTAACCACTGTAACCACTGTAACCACTGTAACCACTGTA 1320
 Db 1320 TGTTTCACACATTTAACTGCTGTAACCACTGTAACCACTGTAACCACTGTAACCACTGTA 1379
 QY 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAGGCTGTTGTTGNAATTAAGGTCG 1380
 Db 1380 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAGGCTGTTGTTGNAATTAAGGTCG 1439
 QY 1381 TAATTCGTATTTCCAGGCTTAGACAAATTAATAGAACGCTTGAATGATGAATCGACTAA 1440
 Db 1440 TAATTCGTATTTCCAGGCTTAGACAAATTAATAGAACGCTTGAATGATGAATCGACTAA 1499
 QY 1441 TAAAGAAATTTGGTAGATGATTTATGGAATTCCTAGACCAATTAACCATCCAGAGCG 1500
 Db 1500 TAAAGAAATTTGGTAGATGATTTATGGAATTCCTAGACCAATTAACCATCCAGAGCG 1559
 QY 1501 ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGAGCAAGTTCGTTATTCCTCAAT 1560
 Db 1560 ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGAGCAAGTTCGTTATTCCTCAAT 1619
 QY 1561 AGCTGATAAGTATACAAAGTTCAGATGGTTACATTTTGTGTAACATGATATATCACTGA 1620
 Db 1620 AGCTGATAAGTATACAAAGTTCAGATGGTTACATTTTGTGTAACATGATATATCACTGA 1679
 QY 1621 TGAAGGAGATGCATATGTAAGGCTCATATGGGCCATAGTCACTGGATGGAAAAGATAG 1680
 Db 1680 TGAAGGAGATGCATATGTAAGGCTCATATGGGCCATAGTCACTGGATGGAAAAGATAG 1739
 QY 1681 CCTTCTGATAGGAAAAAGTTGACGCTCAAGCTATCTAAGAAAAAGGATCCTACC 1740
 Db 1740 CCTTCTGATAGGAAAAAGTTGACGCTCAAGCTATCTAAGAAAAAGGATCCTACC 1799

QY 1741 TCCATCTCCAGACCGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
 Db 1800 TCCATCTCCAGACCGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1859
 QY 1801 CAATCGTGTGAAGGGGAAAAACGAATTCACACTCGTTCCGACTTCCATATATATGGTTGAGCA 1860
 Db 1860 CAATCGTGTGAAGGGGAAAAACGAATTCACACTCGTTCCGACTTCCATATATATGGTTGAGCA 1919
 QY 1861 TACAGTTGAGGTTAAAAACCGTAATTTGATTTATTCCTCATAAGGATCATTAACCAATATAT 1920
 Db 1920 TACAGTTGAGGTTAAAAACCGTAATTTGATTTATTCCTCATAAGGATCATTAACCAATATAT 1979
 QY 1921 TAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGGAAGA 1980
 Db 1980 TAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGGAAGA 2039
 QY 1981 TTTGTTTGGCAGCAATTAAGTACTACGTAGAACACCCCTGACGAAGCTCCACATTTCTAATGA 2040
 Db 2040 TTTGTTTGGCAGCAATTAAGTACTACGTAGAACACCCCTGACGAAGCTCCACATTTCTAATGA 2099
 QY 2041 TGGATGGGCAATGGCCAGTGAAGTGTGTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
 Db 2100 TGGATGGGCAATGGCCAGTGAAGTGTGTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2159
 QY 2101 TAAAGACTTCAAAGCGGATGAAGGCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
 Db 2160 TAAAGACTTCAAAGCGGATGAAGGCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2219
 QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAGAGCAAGTCTTCTTGGGAA 2220
 Db 2220 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAGAGCAAGTCTTCTTGGGAA 2279
 QY 2221 AGTAAAGCAATCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGTGGTTACGAAATAA 2280
 Db 2280 AGTAAAGCAATCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGTGGTTACGAAATAA 2339
 QY 2281 TTTGACTCTTCAAAATATGATAACCAATAGTATCATGCGAGAGCAAGAAATTAATCTTGC 2340
 Db 2340 TTTGACTCTTCAAAATATGATAACCAATAGTATCATGCGAGAGCAAGAAATTAATCTTGC 2399
 QY 2341 GTTGTAAAAAGGAAGTAATCTTCTCATCTGTAAGTAAAGGAAAAATAAAC 2389
 Db 2400 GTTGTAAAAAGGAAGTAATCTTCTCATCTGTAAGTAAAGGAAAAATAAAC 2448

RESULT 3

US-10-412-850-9
 ; Sequence 9, Application US/10412850
 ; Publication No. US20040001836A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-686
 ; CURRENT APPLICATION NUMBER: US/10/412,850
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patentin Ver. 3.0
 ; SEQ ID NO 9
 ; LENGTH: 2451
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(2451)
 ; OTHER INFORMATION: n = a, c, t or g
 US-10-412-850-9


```

1381 QY TAAATCTGATTTCCAGCCTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 1440
1381 Db TAAATCTGATTTCCAGCCTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 1440
1441 QY TAAAGAAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATACCCTCCAGAGCG 1500
1441 Db TAAAGAAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATACCCTCCAGAGCG 1500
1501 QY ACTTGGCAAAACCAATTTCTCAAAATGAGTATATCTGAAGACGAGTTGCTATTTGCTCAAT 1560
1501 Db ACTTGGCAAAACCAATTTCTCAAAATGAGTATATCTGAAGACGAGTTGCTATTTGCTCAAT 1560
1561 QY AGCTCATAGTATACAACTGACAGTGTTCATATTTTGTATGAACATGATATATCAGTGA 1620
1561 Db AGCTCATAGTATACAACTGACAGTGTTCATATTTTGTATGAACATGATATATCAGTGA 1620
1621 QY TGAAGGAGATGATATGTAACGCTCATATGAGGCTATAGTCTAGTGTGAAAAGATAG 1680
1621 Db TGAAGGAGATGATATGTAACGCTCATATGAGGCTATAGTCTAGTGTGAAAAGATAG 1680
1681 QY CTTTCTGATAGGAAAAGTTGAGCTCAAGCTTATCTAAGAAAAGGTTCTCTACC 1740
1681 Db CTTTCTGATAGGAAAAGTTGAGCTCAAGCTTATCTAAGAAAAGGTTCTCTACC 1740
1741 QY TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGACGAGCTATTTA 1800
1741 Db TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGACGAGCTATTTA 1800
1801 QY CAATCGTGTGAAAGGGGAAAAAGCAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1860
1801 Db CAATCGTGTGAAAGGGGAAAAAGCAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1860
1861 QY TACAGTTGAGTTTAAAAAGGTTAATTTGATTTATTCCTCATAGGATCATATCCATATAT 1920
1861 Db TACAGTTGAGTTTAAAAAGGTTAATTTGATTTATTCCTCATAGGATCATATCCATATAT 1920
1921 QY TAAATTTGCTTGTGATGATCACAATCAAGCTCCAAATGGCTATACCTTGGAAAGA 1980
1921 Db TAAATTTGCTTGTGATGATCACAATCAAGCTCCAAATGGCTATACCTTGGAAAGA 1980
1981 QY TTTTGTGCGAGCAATTAAGTACTACGTAGAACACCCCTGACGAACTGCTCAATTTAATGA 2040
1981 Db TTTTGTGCGAGCAATTAAGTACTACGTAGAACACCCCTGACGAACTGCTCAATTTAATGA 2040
2041 QY TGAATGGGCAATGCGAGTGAAGTGTGTAGGCAAGAACACCCCTGACGAACTGCTCAAT 2100
2041 Db TGAATGGGCAATGCGAGTGAAGTGTGTAGGCAAGAACACCCCTGACGAACTGCTCAAT 2100
2101 QY TAAGAACTTCAAGCGGATGAAGGCGAGTGAAGGAAACACCTGCTGAGCGAAAGTCCC 2160
2101 Db TAAGAACTTCAAGCGGATGAAGGCGAGTGAAGGAAACACCTGCTGAGCGAAAGTCCC 2160
2161 QY TCAAGTAGAGACTGAAAAGTAGAAGCCCAATCAAGAGAGCAAGAGTTTGTTCGCA 2220
2161 Db TCAAGTAGAGACTGAAAAGTAGAAGCCCAATCAAGAGAGCAAGAGTTTGTTCGCA 2220
2221 QY AGTAAGGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
2221 Db AGTAAGGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
2281 QY TTTGACTCTTCAAAATGATATGATATGATATGATATGATATGATATGATATGATATG 2340
2281 Db TTTGACTCTTCAAAATGATATGATATGATATGATATGATATGATATGATATGATATG 2340
2341 QY GTTGTAAAAGGAGTAATCTTCACTGTAAGTAAAGGAAAAAATAAAC 2389
2341 Db GTTGTAAAAGGAGTAATCTTCACTGTAAGTAAAGGAAAAAATAAAC 2389

```

RESULT 2
US-10-412-862-9
; Sequence 9, Application US/10412862

```

; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2451)
; OTHER INFORMATION: n = a, c, t or g
US-10-412-862-9

Query Match 100.0%; Score 2389; DB 13; Length 2451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATTAATCGTGTTCCTA 60
Db 60 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATTAATCGTGTTCCTA 119
QY 61 TATAGATGAAAAACAGCGACCCAAAACGGAGAAATTTGACTCCTCGATGAGGTTAGCAA 120
Db 120 TATAGATGAAAAACAGCGACCCAAAACGGAGAAATTTGACTCCTCGATGAGGTTAGCAA 179
QY 121 GGGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db 180 GGGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 239
QY 181 TTCACATGGCGACCACTATCATTTATTAATGATGAGGTTCTTTATGACGCTATCATCAG 240
Db 240 TTCACATGGCGACCACTATCATTTATTAATGATGAGGTTCTTTATGACGCTATCATCAG 299
QY 241 TGAAGAAATTAATCAAGAAATCCAACTATAAGCTAAAAGATGAGGATATGTTAATGA 300
Db 300 TGAAGAAATTAATCAAGAAATCCAACTATAAGCTAAAAGATGAGGATATGTTAATGA 359
QY 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Db 360 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 419
QY 361 TGGCCACGGCGATTAACGTCGCTCAAAAGAGGAAATCAATCGACAAAACAAAGAGCATAG 420
Db 420 TGGCCACGGCGATTAACGTCGCTCAAAAGAGGAAATCAATCGACAAAACAAAGAGCATAG 479
QY 421 TCACAACTCGTGAAGGTGGAACTCCAAAGAAACGATGGTGTGTTGGCTTGGCAAGTTGCGCA 480
Db 480 TCACAACTCGTGAAGGTGGAACTCCAAAGAAACGATGGTGTGTTGGCTTGGCAAGTTGCGCA 539
QY 481 AGGACGCTATCTACTACAGATGATGTTTATATCTTTAATGCTTCGTGATATCATAGAGATAC 540
Db 540 AGGACGCTATCTACTACAGATGATGTTTATATCTTTAATGCTTCGTGATATCATAGAGATAC 599
QY 541 TGTGTATGCTTATATCGTTCCTCATGAGATCATACCATTAATGATGATGATGATGATGAT 600
Db 600 TGTGTATGCTTATATCGTTCCTCATGAGATCATACCATTAATGATGATGATGATGATGAT 659
QY 601 ATCAGCTAGCGAGTTGGCTGCTGCGAAGGCTTCTCTATCTGCTGCGAGGAAATCTGTCAA 660
Db 660 ATCAGCTAGCGAGTTGGCTGCTGCGAAGGCTTCTCTATCTGCTGCGAGGAAATCTGTCAA 719

```


88 19 0.8 7493 15 US-10-187-319-5 GENERAL INFORMA
89 19 0.8 7493 15 US-10-131-510A-5 Sequence 5, Appl
90 19 0.8 91278 17 US-10-322-281-341 Sequence 341, App

ALIGNMENTS

```

RESULT 1
US-09-765-272-55
; Sequence 55, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-765-272-55

Query Match 100.0%; Score 2389; DB 9; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTGGGAGTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 60
Db 1 TTCTTACGAGTGGGAGTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 60
QY 61 TATAGATGGAATAACAGCGACGCAAAACGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
Db 61 TATAGATGGAATAACAGCGACGCAAAACGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
QY 121 GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCAC 180
Db 121 GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCAC 180
QY 181 TTCACATGGCGACCACTATCATTTACAAATGGTAAGGTTCCCTTATGACGCTATCATCAG 240
Db 181 TTCACATGGCGACCACTATCATTTACAAATGGTAAGGTTCCCTTATGACGCTATCATCAG 240
QY 241 TGAAGAATTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300
Db 241 TGAAGAATTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300

```

```

Db 241 TGAAGAATTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300
QY 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Db 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
QY 361 TGCCACGCGGATAACGTCGTACAAAGAGGAAATCAATCGACAAACAAAGAGCATAG 420
Db 361 TGCCACGCGGATAACGTCGTACAAAGAGGAAATCAATCGACAAACAAAGAGCATAG 420
QY 421 TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTGTGCTTGGCCCTTGGCAGCTTCGCA 480
Db 421 TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTGTGCTTGGCCCTTGGCAGCTTCGCA 480
QY 481 AGGACGCTATATACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db 481 AGGACGCTATATACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC 540
QY 541 TGGTGATGCTTATATCGTTCCTCATGGAGATCAATTACCATTAATTCCTAAGAAATGAGTT 600
Db 541 TGGTGATGCTTATATCGTTCCTCATGGAGATCAATTACCATTAATTCCTAAGAAATGAGTT 600
QY 601 ATCAGCTAGCGAGTTGGCTCTCGAGAAAGCTTCTATCTGCTCGAGGAAATCTGTCAA 660
Db 601 ATCAGCTAGCGAGTTGGCTCTCGAGAAAGCTTCTATCTGCTCGAGGAAATCTGTCAA 660
QY 661 TTCAAGAACCTATCGCGGACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
Db 661 TTCAAGAACCTATCGCGGACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
QY 721 TGTAAAGCAATCCAGGACTACAAATACATAACAGCAACAAAGCAACAACTAACAGTCA 780
Db 721 TGTAAAGCAATCCAGGACTACAAATACATAACAGCAACAAAGCAACAACTAACAGTCA 780
QY 781 AGCAAGTCAAAAGTAATGACATTCATGATGCTCTTTGAAACAGCTCTACAAACCTGCTTGG 840
Db 781 AGCAAGTCAAAAGTAATGACATTCATGATGCTCTTTGAAACAGCTCTACAAACCTGCTTGG 840
QY 841 TCAACGACATGTAGAAATCTGATGGCTTGTCTTGTATCCAGCACAAATCAAAAGTCAAC 900
Db 841 TCAACGACATGTAGAAATCTGATGGCTTGTCTTGTATCCAGCACAAATCAAAAGTCAAC 900
QY 901 AGCTAGAGGTGTTGACGTGCCACACGAGATCAATACCACTTCCCTTACTCTCAAT 960
Db 901 AGCTAGAGGTGTTGACGTGCCACACGAGATCAATACCACTTCCCTTACTCTCAAT 960
QY 961 GTCTGAATTGGAAGAACGAATCGCTCGTATTTATTCCTTCTGTTTATCGTTCAAAACATTG 1020
Db 961 GTCTGAATTGGAAGAACGAATCGCTCGTATTTATTCCTTCTGTTTATCGTTCAAAACATTG 1020
QY 1021 GGTACCAGATTCAAGCCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAG 1080
Db 1021 GGTACCAGATTCAAGCCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAG 1080
QY 1081 CCGCAACCTGCAACAAATCTTAAATAGACTCAAAATCTCTTCTGTTAGTCACTGCT 1140
Db 1081 CCGCAACCTGCAACAAATCTTAAATAGACTCAAAATCTCTTCTGTTAGTCACTGCT 1140
QY 1141 ACGAAAGTTGGGAGGAGGATATGTTTCAAGAAAGGAGGATCTCTCGTTATGCTTTTC 1200
Db 1141 ACGAAAGTTGGGAGGAGGATATGTTTCAAGAAAGGAGGATCTCTCGTTATGCTTTTC 1200
QY 1201 GAAAGATTTACCATCTGAACCTGTTTAAATCTTGAAGCAAGGATTTATCAAAACAGAGAG 1260
Db 1201 GAAAGATTTACCATCTGAACCTGTTTAAATCTTGAAGCAAGGATTTATCAAAACAGAGAG 1260
QY 1261 TGTTCACACACTTTAACTGCTTAAAGAGAAATGTTGCTCTCGTCCGACCAAGATTTTA 1320
Db 1261 TGTTCACACACTTTAACTGCTTAAAGAGAAATGTTGCTCTCGTCCGACCAAGATTTTA 1320
QY 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAGGGTGC 1380
Db 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAGGGTGC 1380

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 01:51:39 ; Search time 1910 Seconds
(without alignments)

6340.940 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTACGAGTGGACTGT.....TAAGTACGAAAAATAAAC 2389

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3340653 seqs, 2534783454 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 2389 | 100.0 | 2389 | 9 | US-09-765-272-55 |
| 2 | 2389 | 100.0 | 2451 | 13 | US-10-412-862-9 |
| 3 | 2389 | 100.0 | 2451 | 16 | US-10-412-850-9 |
| 4 | 2389 | 100.0 | 2451 | 16 | US-10-387-783-9 |
| 5 | 2338 | 97.9 | 8195 | 13 | US-10-158-844-94 |
| 6 | 148 | 6.2 | 2523 | 10 | US-09-884-465A-4 |
| 7 | 148 | 6.2 | 2647 | 10 | US-09-884-465A-4 |
| 8 | 53 | 2.2 | 2290 | 9 | US-09-765-272-65 |
| 9 | 53 | 2.2 | 2359 | 13 | US-10-158-844-243 |
| 10 | 53 | 2.2 | 2481 | 10 | US-09-769-787-206 |
| 11 | 53 | 2.2 | 2531 | 13 | US-10-412-862-5 |
| 12 | 53 | 2.2 | 2531 | 13 | US-10-412-862-11 |
| 13 | 53 | 2.2 | 2531 | 16 | US-10-412-850-5 |
| 14 | 53 | 2.2 | 2531 | 16 | US-10-412-850-11 |

| | | | |
|--------|----|----------------------|--------------------|
| 2531 | 16 | US-10-387-783-5 | Sequence 5, Appli |
| 2531 | 16 | US-10-387-783-11 | Sequence 11, Appli |
| 2639 | 10 | US-09-884-465A-5 | Sequence 5, Appli |
| 973 | 13 | US-10-158-844-355 | Sequence 355, App |
| 40 | 9 | US-09-765-272-282 | Sequence 282, App |
| 40 | 13 | US-10-412-862-3 | Sequence 3, Appli |
| 40 | 16 | US-10-412-850-3 | Sequence 3, Appli |
| 40 | 16 | US-10-387-783-3 | Sequence 3, Appli |
| 36 | 13 | US-10-412-862-1 | Sequence 1, Appli |
| 36 | 16 | US-10-412-850-1 | Sequence 1, Appli |
| 36 | 16 | US-10-387-783-1 | Sequence 1, Appli |
| 37 | 9 | US-09-765-272-281 | Sequence 281, App |
| 37 | 13 | US-10-412-862-2 | Sequence 2, Appli |
| 35 | 16 | US-10-412-850-2 | Sequence 2, Appli |
| 35 | 16 | US-10-387-783-2 | Sequence 2, Appli |
| 33 | 10 | US-09-884-465A-45 | Sequence 45, Appli |
| 34 | 10 | US-09-884-465A-47 | Sequence 47, Appli |
| 35 | 10 | US-09-884-465A-174 | Sequence 174, App |
| 1342 | 9 | US-09-765-272-181 | Sequence 181, App |
| 1455 | 10 | US-09-769-787-246 | Sequence 246, App |
| 1455 | 10 | US-09-769-744A-23 | Sequence 23, Appli |
| 1455 | 13 | US-10-412-862-7 | Sequence 7, Appli |
| 1455 | 16 | US-10-412-850-7 | Sequence 7, Appli |
| 1455 | 16 | US-10-387-783-7 | Sequence 7, Appli |
| 2528 | 10 | US-09-884-465A-9 | Sequence 9, Appli |
| 3120 | 10 | US-09-884-465A-1 | Sequence 1, Appli |
| 5048 | 10 | US-09-884-465A-2 | Sequence 2, Appli |
| 6867 | 13 | US-10-158-844-192 | Sequence 192, App |
| 228139 | 13 | US-10-087-192-232 | Sequence 232, App |
| 29 | 10 | US-09-884-465A-44 | Sequence 44, Appli |
| 501 | 16 | US-10-275-933-3 | Sequence 3, Appli |
| 1146 | 10 | US-09-769-736-23 | Sequence 23, Appli |
| 2469 | 10 | US-09-769-736-17 | Sequence 17, Appli |
| 5158 | 16 | US-10-275-933-11 | Sequence 11, Appli |
| 5215 | 10 | US-09-252-088-13 | Sequence 13, Appli |
| 5215 | 16 | US-10-340-792-13 | Sequence 13, Appli |
| 164 | 9 | US-09-983-965-4749 | Sequence 4749, Ap |
| 246 | 17 | US-10-437-963-1529 | Sequence 1529, Ap |
| 427 | 11 | US-09-801-944B-62 | Sequence 62, Appli |
| 483 | 17 | US-10-021-323-6488 | Sequence 6488, Ap |
| 509 | 17 | US-10-021-323-6205 | Sequence 6205, Ap |
| 526 | 17 | US-10-021-323-12910 | Sequence 12910, A |
| 533 | 13 | US-10-027-632-231121 | Sequence 231121, S |
| 533 | 13 | US-10-027-632-231122 | Sequence 231122, S |
| 533 | 13 | US-10-027-632-231123 | Sequence 231123, S |
| 533 | 16 | US-10-027-632-231121 | Sequence 231121, S |
| 533 | 16 | US-10-027-632-231122 | Sequence 231122, S |
| 533 | 16 | US-10-027-632-231123 | Sequence 231123, S |
| 561 | 17 | US-10-021-323-1518 | Sequence 1518, Ap |
| 565 | 17 | US-10-021-323-10473 | Sequence 10473, A |
| 593 | 17 | US-10-021-323-12821 | Sequence 12821, A |
| 594 | 17 | US-10-021-323-9719 | Sequence 9719, Ap |
| 599 | 17 | US-10-021-323-7454 | Sequence 7454, Ap |
| 601 | 17 | US-10-021-323-8160 | Sequence 8160, Ap |
| 658 | 17 | US-10-767-701-25410 | Sequence 25410, A |
| 693 | 16 | US-10-012-697-610 | Sequence 610, App |
| 984 | 17 | US-10-437-963-6825 | Sequence 6825, Ap |
| 1020 | 13 | US-10-282-122A-16763 | Sequence 16763, A |
| 1112 | 13 | US-10-027-632-118539 | Sequence 118539, S |
| 1112 | 13 | US-10-027-632-118540 | Sequence 118540, S |
| 1112 | 16 | US-10-027-632-118539 | Sequence 118539, S |
| 1112 | 16 | US-10-027-632-118540 | Sequence 118540, S |
| 1112 | 17 | US-10-437-963-9121 | Sequence 9121, Ap |
| 1155 | 17 | US-10-437-963-87931 | Sequence 87931, A |
| 1213 | 13 | US-10-027-632-117222 | Sequence 117222, S |
| 1213 | 13 | US-10-027-632-117223 | Sequence 117223, S |
| 1213 | 13 | US-10-027-632-117224 | Sequence 117224, S |
| 1213 | 16 | US-10-027-632-117222 | Sequence 117222, S |
| 1213 | 16 | US-10-027-632-117223 | Sequence 117223, S |
| 1213 | 16 | US-10-027-632-117224 | Sequence 117224, S |
| 1239 | 17 | US-10-767-795-794 | Sequence 794, App |
| 1455 | 10 | US-09-769-736-71 | Sequence 71, Appli |

```

US-08-474-503-5
; Sequence 5, Application US/08474503
; Patent No. 5744446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 1..407
; OTHER INFORMATION: /rpt_type= "terminal"
; OTHER INFORMATION: /note= "5'UTR"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7471..7476
; OTHER INFORMATION: /function= "PolyA_signal"
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 7368..7493
; OTHER INFORMATION: /rpt_type= "terminal"
; OTHER INFORMATION: /note= "3'UTR"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 408..7367
; OTHER INFORMATION: /product= "Coagulation Factor VIII"
; PUBLICATION INFORMATION:
; AUTHORS: Elder, F.
; AUTHORS: Lakich, D.
; AUTHORS: Gitschier, J.
; TITLE: Sequence of the Murine Factor VIII cDNA.
; Patent No. 5744446
; JOURNAL: Genomics
; VOLUME: 16
; PAGES: 374-379
; DATE: 1993
; RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 7476
US-08-474-503-5

```

Query Match

0.8%; Score 19; DB 1; Length 7493;

```

Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2347 AAAAGGAAGTAATCCTTCA 2365
    |||||
Db 206 AAAAGGAAGTAATCCTTCA 224

```

Search completed: October 1, 2004, 03:35:56
Job time : 175 secs

```
Sequence 895, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Rebeschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 895:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1559RP
US-08-998-416-895

Query Match 0.8%; Score 19; DB 3; Length 622;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 TCAAGAACAACTGGGTAC 716
Db 383 TCAAGAACAACTGGGTAC 401

RESULT 24
US-08-212-133A-7
Sequence 7, Application US/08212133A
Patent No. 5663060
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
APPLICANT: Runge, Marshall S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 100 Peachtree Street
CITY: Atlanta
STATE: Georgia

Country: US
ZIP: 30303
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,133A
FILING DATE: March 11, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type= "terminal"
OTHER INFORMATION: /note= "5'UTR"
FEATURE:
NAME/KEY: misc feature
LOCATION: 7471..7476
OTHER INFORMATION: /function= "PolyA_signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type= "terminal"
OTHER INFORMATION: /note= "3'UTR"
FEATURE:
NAME/KEY: misc feature
LOCATION: 408..7367
OTHER INFORMATION: /product= "Coagulation Factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 5663060
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 7476
US-08-212-133A-7

Query Match 0.8%; Score 19; DB 1; Length 7493;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2347 AAAAGGAGTAATCCTTCA 2365
Db 206 AAAAGGAGTAATCCTTCA 224

RESULT 25
```

```
/
/
/ CURRENT APPLICATION DATA:
/   APPLICATION NUMBER: US/09/536,784
/   FILING DATE: 30-Oct-1997
/   CLASSIFICATION: <Unknown>
/
/ PRIOR APPLICATION DATA:
/   APPLICATION NUMBER: 08/961,083
/   FILING DATE: OCT-30-1997
/
/ ATTORNEY/AGENT INFORMATION:
/   NAME: Michelle S. Marks
/   REGISTRATION NUMBER: 41,971
/   REFERENCE/DOCKET NUMBER: PB340P3
/
/ TELECOMMUNICATION INFORMATION:
/   TELEPHONE: (301) 309-8504
/   TELEFAX: (301) 309-8512
/
/ INFORMATION FOR SEQ ID NO: 181:
/   SEQUENCE CHARACTERISTICS:
/     LENGTH: 1342 base pairs
/     TYPE: nucleic acid
/     STRANDEDNESS: double
/     TOPOLOGY: linear
/   SEQUENCE DESCRIPTION: SEQ ID NO: 181:
US-09-536-784-181

Query Match      1.0%; Score 24; DB 4; Length 1342;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCCTCATGGAG 569
Db 525 ATGCTTATATCGTTCCTCATGGAG 548

RESULT 20
US-09-468-656A-7
/ Sequence 7, Application US/09468656A
/ Patent No. 6582706
/ GENERAL INFORMATION:
/   APPLICANT: Johnson, Leslie S.
/   APPLICANT: Adamou, John E.
/   TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
/   TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
/   TITLE OF INVENTION: Motifs
/   FILE REFERENCE: 469201-444
/   CURRENT APPLICATION NUMBER: US/09/468,656A
/   CURRENT FILING DATE: 1999-12-02
/   PRIOR APPLICATION NUMBER: 60/113,048
/   PRIOR FILING DATE: 1998-12-21
/   NUMBER OF SEQ ID NOS: 14
/   SOFTWARE: Patentin Ver. 2.1
/   SEQ ID NO 7
/   LENGTH: 1455
/   TYPE: DNA
/   ORGANISM: Streptococcus pneumoniae
US-09-468-656A-7

Query Match      1.0%; Score 24; DB 4; Length 1455;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCCTCATGGAG 569
Db 596 ATGCTTATATCGTTCCTCATGGAG 619

RESULT 21
US-09-961-527-192/c
/ Sequence 192, Application US/08961527
/ Patent No. 6420135
/ GENERAL INFORMATION:
/   APPLICANT: Charles Kunsch
/   TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
/   NUMBER OF SEQUENCES: 391
/   CORRESPONDENCE ADDRESS:
```

```
/
/
/ ADDRESSEE: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII Text
/
/ CURRENT APPLICATION DATA:
/   APPLICATION NUMBER: US/08/961,527
/   FILING DATE:
/   CLASSIFICATION: 424
/   PRIOR APPLICATION DATA:
/     APPLICATION NUMBER:
/     FILING DATE:
/   ATTORNEY/AGENT INFORMATION:
/     NAME: Brookes, A. Anders
/     REGISTRATION NUMBER: 36,373
/     REFERENCE/DOCKET NUMBER: PB340P1
/     TELECOMMUNICATION INFORMATION:
/       TELEPHONE: (301) 309-8504
/       TELEFAX: (301) 309-8512
/     INFORMATION FOR SEQ ID NO: 192:
/       SEQUENCE CHARACTERISTICS:
/         LENGTH: 6867 base pairs
/         TYPE: nucleic acid
/         STRANDEDNESS: double
/         TOPOLOGY: linear
/   US-08-961-527-192

Query Match      1.0%; Score 24; DB 4; Length 6867;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCCTCATGGAG 569
Db 6190 ATGCTTATATCGTTCCTCATGGAG 6167

RESULT 22
US-09-489-039A-6492
/ Sequence 6492, Application US/09489039A
/ Patent No. 6610836
/ GENERAL INFORMATION:
/   APPLICANT: Gary Breton et. al
/   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
/   TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
/   FILE REFERENCE: 2709.2004001
/   CURRENT APPLICATION NUMBER: US/09/489,039A
/   CURRENT FILING DATE: 2000-01-27
/   PRIOR APPLICATION NUMBER: US 60/117,747
/   PRIOR FILING DATE: 1999-01-29
/   NUMBER OF SEQ ID NOS: 14342
/   SEQ ID NO 6492
/   LENGTH: 768
/   TYPE: DNA
/   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6492

Query Match      0.8%; Score 20; DB 4; Length 768;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2368 TGTAAAGTAGGAAAAATAA 2387
Db 183 TGTAAAGTAGGAAAAATAA 202

RESULT 23
US-08-998-416-895
```

STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 281:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 281:
US-09-536-784-281

Query Match 1.1%; Score 27; DB 4; Length 37;
Best Local Similarity 100.0%; Pred. No. 0.00093; Indels 0;
Matches 27; Conservative 0; Mismatches 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGC 27
DB 11 TTCTTACGAGTTGGGACTGTATCAAGC 37

RESULT 17
US-09-468-656A-2
Sequence 2, Application US/09468656A
Patent No. 6582706
GENERAL INFORMATION:
APPLICANT: Johnson, Leslie S.
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
TITLE OF INVENTION: Motifs
FILE REFERENCE: 469201-444
CURRENT APPLICATION NUMBER: US/09/468,656A
CURRENT FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: 60/113,048
PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 35
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Forward primer
OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-09-468-656A-2

Query Match 1.1%; Score 26; DB 4; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.003; Indels 0;
Matches 26; Conservative 0; Mismatches 0; Gaps 0;

QY 16 ACTGTATCAAGCTAGACGGTTAAGG 41
DB 10 ACTGTATCAAGCTAGACGGTTAAGG 35

RESULT 18
US-08-961-083-181
Sequence 181, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 181:
SEQUENCE CHARACTERISTICS:
LENGTH: 1342 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-181

Query Match 1.0%; Score 24; DB 3; Length 1342;
Best Local Similarity 100.0%; Pred. No. 0.037; Indels 0;
Matches 24; Conservative 0; Mismatches 0; Gaps 0;

QY 546 ATGCTTATATCGTTCTCTCATGGAG 569
DB 525 ATGCTTATATCGTTCTCTCATGGAG 548

RESULT 19
US-09-536-784-181
Sequence 181, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

```
;
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 282:
US-09-536-784-282

Query Match      1.3%; Score 30; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 CTTTCATCTGTAAGTAAGGAAAAATAAAC 2389
Db 40 CTTTCATCTGTAAGTAAGGAAAAATAAAC 11

RESULT 13
US-09-468-656A-3/c
; Sequence 3, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR FILING DATE: 1999-12-02
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Reverse primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-09-468-656A-3

Query Match      1.3%; Score 30; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 CTTTCATCTGTAAGTAAGGAAAAATAAAC 2389
Db 40 CTTTCATCTGTAAGTAAGGAAAAATAAAC 11

RESULT 14
US-09-468-656A-1
; Sequence 1, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR FILING DATE: 1999-12-02
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 36
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Forward primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-09-468-656A-1

;
; Query Match      1.1%; Score 27; DB 3; Length 37;
; Best Local Similarity 100.0%; Pred. No. 0.00093;
; Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGC 27
Db 10 TTCTTACGAGTTGGGACTGTATCAAGC 36

RESULT 15
US-08-961-083-281
; Sequence 281, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 281:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-281

Query Match      1.1%; Score 27; DB 3; Length 37;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGC 27
Db 11 TTCTTACGAGTTGGGACTGTATCAAGC 37

RESULT 16
US-09-536-784-281
; Sequence 281, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; 
```

```

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 282:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-282

Query Match 1.3%; Score 30; DB 3; Length 40;
Best Local similarity 100.0%; Pred. No. 2.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 CCTTCATCTGTAAGTAAGGAAAAATAAAC 2389
Db ||||||||||||||||||||||||||||
40 CCTTCATCTGTAAGTAAGGAAAAATAAAC 11

RESULT 12
US-09-536-784-282/c
Sequence 282, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 282:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

```


MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 2290 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 65:
US-09-536-784-65

Query Match 2.2%; Score 53; DB 4; Length 2290;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 1035
|||||
DB 944 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 996
|||||

RESULT 7
US-09-536-784-65
; Sequence 243, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 243:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-243

Query Match 2.2%; Score 53; DB 4; Length 2359;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 1035
|||||
DB 1879 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 1931
|||||

RESULT 8
US-09-468-656A-5
; Sequence 5, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-5

Query Match 2.2%; Score 53; DB 4; Length 2531;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 1035
|||||
DB 1003 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 1055
|||||

RESULT 9
US-09-468-656A-11
; Sequence 11, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-11

Query Match 2.2%; Score 53; DB 4; Length 2531;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4253 GAAAGATTACCATCTGAACTGTAAATCTTGAAGCAAGTTATCAAAACAAGAG 4312
1261 TGTTTCCACACTTTAACTGCTAAAGAAAGAAATGTTGCTCCTCGTGAACCAAGAAATTTTA 1320
4313 TGTTTCCACACTTTAACTGCTAAAGAAAGAAATGTTGCTCCTCGTGAACCAAGAAATTTTA 4372
1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTGNAATATAGGGTCG 1380
4373 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTGNAATATAGGGTCG 4432
1391 TAAATCTCATTTCCAGCCTTAGACAAATATTAGAACGCTTGAATGATGATCACTAA 1440
4433 TAAATCTCATTTCCAGCCTTAGACAAATATTAGAACGCTTGAATGATGATCACTAA 4492
1441 TAAAGAAAAATGGTAGATGATTTATTGGCAFTCTAGACCAATACCATCCAGAGGG 1500
4493 TAAAGAAAAATGGTAGATGATTTATTGGCAFTCTAGACCAATACCATCCAGAGGG 4552
1501 ACTTGGCAAAACCAATCTCAATTTAGATGATGATGATGATGATGATGATGATGAT 1560
4553 ACTTGGCAAAACCAATCTCAATTTAGATGATGATGATGATGATGATGATGATGAT 4612
1561 AGCTGATAGTATACAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
4613 AGCTGATAGTATACAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4672
1621 TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCTAGATGGAAGATAG 1680
4673 TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCTAGATGGAAGATAG 4732
1681 CTTTCTGATAGGAAAAAGTTGACGCTCAAGCTTATTAAGAAAAAGGTTATCTACC 1740
4733 CTTTCTGATAGGAAAAAGTTGACGCTCAAGCTTATTAAGAAAAAGGTTATCTACC 4792
1741 TCCATCTCAGAGCAGATGTTAAAGCAATCACTGAGAGATGAGCAGCACTATTTA 1800
4793 TCCATCTCAGAGCAGATGTTAAAGCAATCACTGAGAGATGAGCAGCACTATTTA 4852
1801 CAATCTGTGAAAGGGGAAAAACGAATTCACCTGCTGATGATGATGATGATGATGATGAT 1860
4853 CAATCTGTGAAAGGGGAAAAACGAATTCACCTGCTGATGATGATGATGATGATGATGAT 4912
1861 TACAGTTGAGTTAAAGCGTAATTTGATGATGATGATGATGATGATGATGATGATGAT 1920
4913 TACAGTTGAGTTAAAGCGTAATTTGATGATGATGATGATGATGATGATGATGATGAT 4972
1921 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
4973 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5032
1981 TTTGTTTGGCAGCATTAAGTACTACGTPAGAACACCTTGAACAGCTCCACATTTAATGA 2040
5033 TTTGTTTGGCAGCATTAAGTACTACGTPAGAACACCTTGAACAGCTCCACATTTAATGA 5092
2041 TGGATGGGGCAATGCCAGTACGATGTTAGGCAAGAACACACAGTGAAGATCCAAA 2100
5093 TGGATGGGGCAATGCCAGTACGATGTTAGGCAAGAACACACAGTGAAGATCCAAA 5152
2101 TAAAGACTTCAAGCGGATGAAGAGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
5153 TAAAGACTTCAAGCGGATGAAGAGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 5212
2161 TCAAGTAGAGACTGAAAGAGTAGAAGCCCACTCAAGAGCAGAGATTTGCTTGGCAA 2220
5213 TCAAGTAGAGACTGAAAGAGTAGAAGCCCACTCAAGAGCAGAGATTTGCTTGGCAA 5272
2221 AGTAACGATTTCTAGTCTGAAGCCCACTGCAACAGAACTCTAGCTGTTTACGAATAA 2280
5273 AGTAACGATTTCTAGTCTGAAGCCCACTGCAACAGAACTCTAGCTGTTTACGAATAA 5332
2281 TTTGACTCTTCAAAATTTAGGATACAAATAGTATCATGCGAGAGCAGAAAAATTTCTTGC 2340
5333 TTTGACTCTTCAAAATTTAGGATACAAATAGTATCATGCGAGAGCAGAAAAATTTCTTGC 5392

2341 GTTGTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAAATAAAC 2389
5393 GTTGTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAAATAAAC 5441

RESULT 5
US-08-961-083-65
; Sequence 65, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-083-65

Query Match 2.2%; Score 53; DB 3; Length 2290;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

983 GCTCGTATTATTCCTTCCTGTTTCAAAACCATTTGGTACCAGATTCAG 1035
944 GCTCGTATTATTCCTTCCTGTTTCAAAACCATTTGGTACCAGATTCAG 996

RESULT 6
US-09-536-784-65
; Sequence 65, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:

QY 2161 TCAAGTAGAGACTGAAAGAGTAGAGCCCAACTCAAAGAGCAGAGTTTTCCTTGGCA 2220
Db 2220 TCAAGTAGAGACTGAAAGAGTAGAGCCCAACTCAAAGAGCAGAGTTTTCCTTGGCA 2279
QY 2221 AGTAACGGATCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
Db 2280 AGTAACGGATCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2339
QY 2281 TTTGACTCTTCAAAATATGGAATAAATAAGTATCATGCGAGAGCAGAAAAATTTACTTGC 2340
Db 2340 TTTGACTCTTCAAAATATGGAATAAATAAGTATCATGCGAGAGCAGAAAAATTTACTTGC 2399
QY 2341 GTTGTAAAGGAAGTAATCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389
Db 2400 GTTGTAAAGGAAGTAATCTTCTATCTGTAAGTAAGGAAAAATAAAC 2448

RESULT 4
US-08-961-527-94
; Sequence 94, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-94

Query Match 97.9%; Score 2338; DB 4; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGATCAAGCTAGAACGGTTAAGGAAAAATAATCGTTTCCTA 60
Db 3053 TTCTTACGAGTTGGAGCTGATCAAGCTAGAACGGTTAAGGAAAAATAATCGTTTCCTA 3112
QY 61 TATAGATGGAACAAAGCGCAGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120
Db 3113 TATAGATGGAACAAAGCGCAGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA 3172
QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAAGATAACAGACCAAGGCTATGTCAC 180

Db 3173 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAAGATAACAGACCAAGGCTATGTCAC 3232
QY 181 TTCAATCGCGACCACTATCATTATTACAAATGGTGAAGTTTCCCTTATGAGGCTATCATCAG 240
Db 3233 TTCAATCGCGACCACTATCATTATTACAAATGGTGAAGTTTCCCTTATGAGGCTATCATCAG 3292
QY 241 TGAAGAATTAATCTGATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAAATGA 300
Db 3293 TGAAGAATTAATCTGATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAAATGA 3352
QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Db 3353 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 3412
QY 361 TGCCACACGGGATAACGTCCTGACAAAGAGGAATCAATCGACAAACCAAGAGCATAG 420
Db 3413 TGCCACACGGGATAACGTCCTGACAAAGAGGAATCAATCGACAAACCAAGAGCATAG 3472
QY 421 TCAACATCGTGAAGGTGGAACCTCCAAGAAACGATGGTGCTGTTGCTTGGCACCGTTCCGA 480
Db 3473 TCAACATCGTGAAGGTGGAACCTCCAAGAAACGATGGTGCTGTTGCTTGGCACCGTTCCGA 3532
QY 481 AGGACGCTATACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db 3533 AGGACGCTATACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC 3592
QY 541 TGGTGAATCTTATATCGTTCTCATGAGATCATTAAGTATGATTTCTTAAAGATGAGTT 600
Db 3593 TGGTGAATCTTATATCGTTCTCATGAGATCATTAAGTATGATTTCTTAAAGATGAGTT 3652
QY 601 ATCAGCTAGCGAGTTGGCTGTCAGAAAGCTTCTATCTGGTTCGAGAAATCTGTCAA 660
Db 3653 ATCAGCTAGCGAGTTGGCTGTCAGAAAGCTTCTATCTGGTTCGAGAAATCTGTCAA 3712
QY 661 TTCAAGAAACCTATCGCGACAAATAGCAGTAACACTTTCAAGAACAACTTGGGTACCTTC 720
Db 3713 TTCAAGAAACCTATCGCGACAAATAGCAGTAACACTTTCAAGAACAACTTGGGTACCTTC 3772
QY 721 TGTAAGCAATCCAGGAATCAAAATCTAAACAAAGCAACAAAGCAAACTTCAAGTACCTTC 780
Db 3773 TGTAAGCAATCCAGGAATCAAAATCTAAACAAAGCAACAAAGCAAACTTCAAGTACCTTC 3832
QY 781 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAACACAGCTCTACAAATCGCTTTGAG 840
Db 3833 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAACACAGCTCTACAAATCGCTTTGAG 3892
QY 841 TCAACGACATGTAGAATCTGATGCGCTTGTGTTGATCCAGCACAATCAAGTTCGAAC 900
Db 3893 TCAACGACATGTAGAATCTGATGCGCTTGTGTTGATCCAGCACAATCAAGTTCGAAC 3952
QY 901 AGCTAGAGTTGTTGCGAGTGCACACGAGATCATTAACCACTTCCCTTACTCTCAAT 960
Db 3953 AGCTAGAGTTGTTGCGAGTGCACACGAGATCATTAACCACTTCCCTTACTCTCAAT 4012
QY 961 GTCTGAATTTGGAAGAACGAATCGCTCGTATATTCCCTTGGTTATCGTTTCAAAACCATG 1020
Db 4013 GTCTGAATTTGGAAGAACGAATCGCTCGTATATTCCCTTGGTTATCGTTTCAAAACCATG 4072
QY 1021 GGTACCAGATTTCAAGCCGAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAG 1080
Db 4073 GGTACCAGATTTCAAGCCGAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAG 4132
QY 1081 CCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGCT 1140
Db 4133 CCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGCT 4192
QY 1141 AGCAAAAGTTGGGAGAGGATATGTTATCGAAGAAAGGCACTCTCTGTTATGCTTTTC 1200
Db 4193 AGCAAAAGTTGGGAGAGGATATGTTATCGAAGAAAGGCACTCTCTGTTATGCTTTTC 4252
QY 1201 GAAAGATTTACCCTGTAACCTGTTAAATCTTTGAGCAAGTATCAAAACAGAGAG 1260

| | | Best Local Similarity | 100.0%; | Pred. No. 0; | Mismatches | 0; | Indels | 0; | Gaps | 0; |
|----|------|--|---------|--------------|------------|----|--------|----|------|----|
| | | Matches 2389; Conservative 0; | | | | | | | | |
| Qy | 1 | TTCCTACAGTTGGGACTGTATCAAGCTAGAACCGGTTAAGGAAATAATATCGTGTTCCTA 60 | | | | | | | | |
| Db | 60 | TTCCTACAGTTGGGACTGTATCAAGCTAGAACCGGTTAAGGAAATAATATCGTGTTCCTA 119 | | | | | | | | |
| Qy | 61 | TATAGATCGAAAAAAGCGAGCGCAAAAAACGGAGAAATTGACTCCTGATGAGGTTAGCAA 120 | | | | | | | | |
| Db | 120 | TATAGATCGAAAAAAGCGAGCGCAAAAAACGGAGAAATTGACTCCTGATGAGGTTAGCAA 179 | | | | | | | | |
| Qy | 121 | CGGTGAAGAAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 180 | | | | | | | | |
| Db | 180 | CGGTGAAGAAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 239 | | | | | | | | |
| Qy | 181 | TTCACATGGCGACCACTATCAATATTACAATGGTAAGGTTCTTATGAGCGCTATCATCAG 240 | | | | | | | | |
| Db | 240 | TTCACATGGCGACCACTATCAATATTACAATGGTAAGGTTCTTATGAGCGCTATCATCAG 299 | | | | | | | | |
| Qy | 241 | TGAAGAAATTACTCATGAAGATCCAAACTATTAAGCTTAAGATGAGGATTTGTTAATGA 300 | | | | | | | | |
| Db | 300 | TGAAGAAATTACTCATGAAGATCCAAACTATTAAGCTTAAGATGAGGATTTGTTAATGA 359 | | | | | | | | |
| Qy | 301 | GGTCAAGGTTGATATGTTATCAAGGTAGATCGAAATATCTATGTTTACCTTAAAGGATGC 360 | | | | | | | | |
| Db | 360 | GGTCAAGGTTGATATGTTATCAAGGTAGATCGAAATATCTATGTTTACCTTAAAGGATGC 419 | | | | | | | | |
| Qy | 361 | TGCCCAACGCGGATACGCTCCGTCAAAAGAGGAAATCAATCGACAAAAACAAGAGATAG 420 | | | | | | | | |
| Db | 420 | TGCCCAACGCGGATACGCTCCGTCAAAAGAGGAAATCAATCGACAAAAACAAGAGATAG 479 | | | | | | | | |
| Qy | 421 | TCAACATCGTGAAGTGGAACTCCAGAAACGATGCTGTTGCTTCGCGAGTTCGCA 480 | | | | | | | | |
| Db | 480 | TCAACATCGTGAAGTGGAACTCCAGAAACGATGCTGTTGCTTCGCGAGTTCGCA 539 | | | | | | | | |
| Qy | 481 | AGGACGCTACTACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC 540 | | | | | | | | |
| Db | 540 | AGGACGCTACTACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC 599 | | | | | | | | |
| Qy | 541 | TGTGTATGCTTATATCGTTCTCATGAGATCAATACCAATTAATTCCTTAAGAAATGAGTT 600 | | | | | | | | |
| Db | 600 | TGTGTATGCTTATATCGTTCTCATGAGATCAATACCAATTAATTCCTTAAGAAATGAGTT 659 | | | | | | | | |
| Qy | 601 | ATCAGTACGAGTGGCTGCTCAGAGACCTTCTATCTGCTCGAGGAATCTGCAAA 660 | | | | | | | | |
| Db | 660 | ATCAGTACGAGTGGCTGCTCAGAGACCTTCTATCTGCTCGAGGAATCTGCAAA 719 | | | | | | | | |
| Qy | 661 | TTCAGAACTTATCGCGCAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720 | | | | | | | | |
| Db | 720 | TTCAGAACTTATCGCGCAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 779 | | | | | | | | |
| Qy | 721 | TGTAAAGCAATCCAGGAATCAAAATCTAAACAAGCAACAAGCAAACTAACAGTCA 780 | | | | | | | | |
| Db | 780 | TGTAAAGCAATCCAGGAATCAAAATCTAAACAAGCAACAAGCAAACTAACAGTCA 839 | | | | | | | | |
| Qy | 781 | AGCAAGTCAAGTAATGACATTTGATGCTCTTTGAACAGCTCTACAAACTGCTTTGAG 840 | | | | | | | | |
| Db | 840 | AGCAAGTCAAGTAATGACATTTGATGCTCTTTGAACAGCTCTACAAACTGCTTTGAG 899 | | | | | | | | |
| Qy | 841 | TCAACGACATGTAGAATCTGATGCTTGTCTTTGATCCAGCAAAATCAAGTCAAGC 900 | | | | | | | | |
| Db | 900 | TCAACGACATGTAGAATCTGATGCTTGTCTTTGATCCAGCAAAATCAAGTCAAGC 959 | | | | | | | | |
| Qy | 901 | AGCTAGAGTGTGTCAGTGCACACGAGATCAATACCACTTATCGCTTACTCTCAAT 960 | | | | | | | | |
| Db | 960 | AGCTAGAGTGTGTCAGTGCACACGAGATCAATACCACTTATCGCTTACTCTCAAT 1019 | | | | | | | | |
| Qy | 961 | GTCTGAATTCGAAGCAAGATCGCTCTGTTATTTTCCCTTCTGTTATCGTTCAAAACCATG 1020 | | | | | | | | |
| Db | 1020 | GTCTGAATTCGAAGCAAGATCGCTCTGTTATTTTCCCTTCTGTTATCGTTCAAAACCATG 1079 | | | | | | | | |
| Qy | 1021 | GGTACCAGATTCAGGCGCAGAACCAACCAAGTCCCAACCGGACTCCGGAACTAGTCCAGG 1080 | | | | | | | | |

| | | |
|----|------|---|
| Db | 1080 | GGTACCAGATTCAGGCGCAGAACCAACCAAGTCCCAACCGGACTCCGAACTAGTCCAGG 1139 |
| Qy | 1081 | CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGCT 1140 |
| Db | 1140 | CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGCT 1199 |
| Qy | 1141 | ACGAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGCACTCTCTGTTATGTTGCTTGC 1200 |
| Db | 1200 | ACGAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGCACTCTCTGTTATGTTGCTTGC 1259 |
| Qy | 1201 | GAAAGATTTACCATCTGAACTGTTAAATCTTGAAGCAAGTCTTATCAAAAACAAGAGAG 1260 |
| Db | 1260 | GAAAGATTTACCATCTGAACTGTTAAATCTTGAAGCAAGTCTTATCAAAAACAAGAGAG 1319 |
| Qy | 1261 | TGTTTTACACACTTTAACTGCTTAAAAAGAAAAATGTTGCTCTCGTACCAAGAAATTTTA 1320 |
| Db | 1320 | TGTTTTACACACTTTAACTGCTTAAAAAGAAAAATGTTGCTCTCGTACCAAGAAATTTTA 1379 |
| Qy | 1321 | TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGCTG 1380 |
| Db | 1380 | TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGCTG 1439 |
| Qy | 1381 | TAATTTCTGANTTTCCAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1440 |
| Db | 1440 | TAATTTCTGANTTTCCAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1499 |
| Qy | 1441 | TAAAGAAAAATTTGGTAGATGATTTATTGTCATTCTTAGCACCAATTAACCATCCAGAGCG 1500 |
| Db | 1500 | TAAAGAAAAATTTGGTAGATGATTTATTGTCATTCTTAGCACCAATTAACCATCCAGAGCG 1559 |
| Qy | 1501 | ACTTGGCAAAACCAATTTCTCAAATTTGAGTATATCTGAAGAGCAAGTTCTGTTCAAT 1560 |
| Db | 1560 | ACTTGGCAAAACCAATTTCTCAAATTTGAGTATATCTGAAGAGCAAGTTCTGTTCAAT 1619 |
| Qy | 1561 | AGCTGATAGTATACAACTGAGTGTACATTTTGTGATGAACATGATATAATCACTGTA 1620 |
| Db | 1620 | AGCTGATAGTATACAACTGAGTGTACATTTTGTGATGAACATGATATAATCACTGTA 1679 |
| Qy | 1621 | TGAAGAGATGTCATATGTAACGCTCATATGCGCCATAGTCTACTGGATTGGAAAAAGATG 1680 |
| Db | 1680 | TGAAGAGATGTCATATGTAACGCTCATATGCGCCATAGTCTACTGGATTGGAAAAAGATG 1739 |
| Qy | 1681 | CTTTTCTGATTAAGAAAAAGTTGACGCTCAAGCTTATCTAAAGAAAAAGGATCTCTACC 1740 |
| Db | 1740 | CTTTTCTGATTAAGAAAAAGTTGACGCTCAAGCTTATCTAAAGAAAAAGGATCTCTACC 1799 |
| Qy | 1741 | TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAAGATAGTGCAGCAGCTATTTA 1800 |
| Db | 1800 | TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAAGATAGTGCAGCAGCTATTTA 1859 |
| Qy | 1801 | CAATCGTGTGAAAGGGAAAAAGAAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1860 |
| Db | 1860 | CAATCGTGTGAAAGGGAAAAAGAAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1919 |
| Qy | 1861 | TACAGTTGAGGTTAAAAACCGGTAATTTGATTTTCTCATAGGATCAITTACCATATAT 1920 |
| Db | 1920 | TACAGTTGAGGTTAAAAACCGGTAATTTGATTTTCTCATAGGATCAITTACCATATAT 1979 |
| Qy | 1921 | TAAATTTGCTTGTGTTTGTATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980 |
| Db | 1980 | TAAATTTGCTTGTGTTTGTATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGA 2039 |
| Qy | 1981 | TTTGTGTTGCGACGATTAAGTACTACGTPAGAACCCCTGACGAACGCTCCACATTTCAATGA 2040 |
| Db | 2040 | TTTGTGTTGCGACGATTAAGTACTACGTPAGAACCCCTGACGAACGCTCCACATTTCAATGA 2099 |
| Qy | 2041 | TGAGTGGGCAATGCCAGTGAAGCTGTTAGGCAAGAAACACCACTGAGTGAAGTCCAAA 2100 |
| Db | 2100 | TGAGTGGGCAATGCCAGTGAAGCTGTTAGGCAAGAAACACCACTGAGTGAAGTCCAAA 2159 |
| Qy | 2101 | TAAAGCTTCCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160 |
| Db | 2160 | TAAAGCTTCCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2219 |

1561 AGCTGATAGTATACAACGTCAGATGGTTACATTTTGTGATGAATGATATATCAATCAAGTCA 1620
1621 TGAAGGAGATGATATGTAAGCCCTCATATGGGCCATAGTCTACTGGATTGGAAGATAG 1680
1621 TGAAGGAGATGATATGTAAGCCCTCATATGGGCCATAGTCTACTGGATTGGAAGATAG 1680
1681 CCTTCTCATAGGAAAAGTTGACGCTCAAGCCTATATCTAAGAAAAGGATCTCTACC 1740
1681 CCTTCTCATAGGAAAAGTTGACGCTCAAGCCTATATCTAAGAAAAGGATCTCTACC 1740
1741 TCCATCTCCAGACGAGATGTTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
1741 TCCATCTCCAGACGAGATGTTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
1801 CAATCGTGTGAAGGGGAAAACGAATTCACCTGCTGGTACATTCATATATGTTGAGCA 1860
1801 CAATCGTGTGAAGGGGAAAACGAATTCACCTGCTGGTACATTCATATATGTTGAGCA 1860
1861 TACAGTTGAGGTTTAAAGCGTAATTTGATTATCTCTAAGGATCATTTACCAATAT 1920
1861 TACAGTTGAGGTTTAAAGCGTAATTTGATTATCTCTAAGGATCATTTACCAATAT 1920
1921 TAAATTTGCTTGGTTGATGATCAACATPACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
1921 TAAATTTGCTTGGTTGATGATCAACATPACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
1981 TTTGTTGGGAGATTAAGTACTAGTAGAACACCTGACCAAGCTCCCAATCTTAATGA 2040
1981 TTTGTTGGGAGATTAAGTACTAGTAGAACACCTGACCAAGCTCCCAATCTTAATGA 2040
2041 TGGATGGGCAATGCGAGTACATGTTAGGCAAGAACCCACAGTGAAGATCCAAA 2100
2041 TGGATGGGCAATGCGAGTACATGTTAGGCAAGAACCCACAGTGAAGATCCAAA 2100
2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
2161 TCAAGTAGAGATGAAAAGTAGAGCCCACTCAAGAGCAGAGTTTTCCTTGGAA 2220
2161 TCAAGTAGAGATGAAAAGTAGAGCCCACTCAAGAGCAGAGTTTTCCTTGGAA 2220
2221 AGTAACGGATCTAGTCTCAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
2221 AGTAACGGATCTAGTCTCAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
2281 TTTGACTCTTCAAATTTATGATAAACAATAGTATGCGAGAGCAAGAAAATTTACTTGC 2340
2281 TTTGACTCTTCAAATTTATGATAAACAATAGTATGCGAGAGCAAGAAAATTTACTTGC 2340
2341 GTTGTAAAGGAAGTAATCTCTTCACTCTGTAAGTAAAGGAAAATAAAC 2389
2341 GTTGTAAAGGAAGTAATCTCTTCACTCTGTAAGTAAAGGAAAATAAAC 2389

RESULT 2

US-09-536-784-55

Sequence 55, Application US/09536784
Patent No. 6573082

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 2389 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-536-784-55

Query Match 100.0%; Score 2389; DB 4; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTTA 60
Db 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTTA 60
QY 61 TATAGATGAAAAACAAGCGACGCAAAAACGGAGATTTTGACTTCCTGATGAGGTTAGCAA 120
Db 61 TATAGATGAAAAACAAGCGACGCAAAAACGGAGATTTTGACTTCCTGATGAGGTTAGCAA 120
QY 121 GGTGTAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db 121 GGTGTAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
QY 181 TTCAATGCGGACCACTATCATATTATTAATGTAAGGTTCTTATGAGCCTATCATCAG 240
Db 181 TTCAATGCGGACCACTATCATATTATTAATGTAAGGTTCTTATGAGCCTATCATCAG 240
QY 241 TGAAGAAATTACTCATCAAGATCCAACTATAAGCTAAAGATGAGGATTTGTTAATGA 300
Db 241 TGAAGAAATTACTCATCAAGATCCAACTATAAGCTAAAGATGAGGATTTGTTAATGA 300
QY 301 GGTCAAGGTTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Db 301 GGTCAAGGTTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
QY 361 TGCCCAACGCGGATAACGTCCTGACAAAAGAGAAATCAATCGCAAAAAACAAGAGCATAG 420
Db 361 TGCCCAACGCGGATAACGTCCTGACAAAAGAGAAATCAATCGCAAAAAACAAGAGCATAG 420
QY 421 TCACATCGTGAAGTGGAACTCCAAAGAACGATGGTGTGCTTGGCTTGGCAGCTTCGCA 480
Db 421 TCACATCGTGAAGTGGAACTCCAAAGAACGATGGTGTGCTTGGCTTGGCAGCTTCGCA 480
QY 481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db 481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
QY 541 TGGTATGCTTATATCGTTCTTCATGAGATCATTTACCAATACATTTCTTAAGAAATGAGTT 600
Db 541 TGGTATGCTTATATCGTTCTTCATGAGATCATTTACCAATACATTTCTTAAGAAATGAGTT 600
QY 601 ATCAGCTAGGAGTTGGCTGCTGCAGAACCTTCTCTATCTGGTGCAGGAAATCTGTCAAA 660
Db 601 ATCAGCTAGGAGTTGGCTGCTGCAGAACCTTCTCTATCTGGTGCAGGAAATCTGTCAAA 660

```

;
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-961-083-55
;
; Query Match 100.0%; Score 2389; DB 3; Length 2389;
; Best Local Similarity 100.0%; Pred. No. 0;
; Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 TTCTTAGAGTTGGAGCTGATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA 60
; DB 1 TTCTTAGAGTTGGAGCTGATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA 60
;
; QY 61 TATAGATGGAAACACGACGCGGAAACCGAGAAATTTGACTCCTGATGAGGTAGCAA 120
; DB 61 TATAGATGGAAACACGACGCGGAAACCGAGAAATTTGACTCCTGATGAGGTAGCAA 120
;
; QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
; DB 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
;
; QY 181 TTCAATGGCGACACATATCAATTTAATGTAAGTTTCCCTTATGACCTATCATCAG 240
; DB 181 TTCAATGGCGACACATATCAATTTAATGTAAGTTTCCCTTATGACCTATCATCAG 240
;
; QY 241 TGAAGAAATTAATCAATGAAGATCCAACTATAGCTTAAAGATGAGGATTTGTTAATGA 300
; DB 241 TGAAGAAATTAATCAATGAAGATCCAACTATAGCTTAAAGATGAGGATTTGTTAATGA 300
;
; QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360
; DB 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360
;
; QY 361 TGCCCGACGCGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAACCAAGAGCATAG 420
; DB 361 TGCCCGACGCGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAACCAAGAGCATAG 420
;
; QY 421 TCAACATCGTGAAGGTGGAACTCAAGAAACGATGGTGTGCTGCTTGGCGACGTTGCGCA 480
; DB 421 TCAACATCGTGAAGGTGGAACTCAAGAAACGATGGTGTGCTGCTTGGCGACGTTGCGCA 480

```

```

;
; QY 481 AGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
; DB 481 AGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
;
; QY 541 TGGTGATGCTTATATCTGCTCATGAGATCATACCATTACATTCCTTAAGATGAGTT 600
; DB 541 TGGTGATGCTTATATCTGCTCATGAGATCATACCATTACATTCCTTAAGATGAGTT 600
;
; QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCTATCTGGTCGAGGAAATCTGTCAA 660
; DB 601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCTATCTGGTCGAGGAAATCTGTCAA 660
;
; QY 661 TTCAAGAACCTATCGCGGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
; DB 661 TTCAAGAACCTATCGCGGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
;
; QY 721 TGTAAAGCAATCCAGGAACATACTAACAGCAACAAAGCAACAAAGCAACACTTACAGTCA 780
; DB 721 TGTAAAGCAATCCAGGAACATACTAACAGCAACAAAGCAACAAAGCAACACTTACAGTCA 780
;
; QY 781 AGCAAGTCAAAAGTAATGACATTTAGTCTCTTGAAGCAAGCTCTACAACTGCTTTGAG 840
; DB 781 AGCAAGTCAAAAGTAATGACATTTAGTCTCTTGAAGCAAGCTCTACAACTGCTTTGAG 840
;
; QY 841 TCAACGACATGTAGAATCTGATGCGCTTGTGATCCAGACAAATCAACAGTCAAC 900
; DB 841 TCAACGACATGTAGAATCTGATGCGCTTGTGATCCAGACAAATCAACAGTCAAC 900
;
; QY 901 AGCTAGAGGTGTTGCGAGTGCACACGAGATCATTAACCACTTCTCTTCTTCTCTCAAA 960
; DB 901 AGCTAGAGGTGTTGCGAGTGCACACGAGATCATTAACCACTTCTCTTCTTCTCTCAAA 960
;
; QY 961 GTCTGAATTGGAAGAACGAAATCGCTCGTATTTATTCCTTCTGTTTCAAAACCATG 1020
; DB 961 GTCTGAATTGGAAGAACGAAATCGCTCGTATTTATTCCTTCTGTTTCAAAACCATG 1020
;
; QY 1021 GGTACACAGATTCAGGCGCAGAACCAAGTCCACAAACGACTCCGGAACCTAGTCCAGG 1080
; DB 1021 GGTACACAGATTCAGGCGCAGAACCAAGTCCACAAACGACTCCGGAACCTAGTCCAGG 1080
;
; QY 1081 CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTCTTCTTCTGTTAGTCAAGTGGT 1140
; DB 1081 CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTCTTCTTCTGTTAGTCAAGTGGT 1140
;
; QY 1141 ACAGAAAGTTGGGAGGATATGATTTGGAAGAAAGGGCATCTCTCGTTATGCTTTTGC 1200
; DB 1141 ACAGAAAGTTGGGAGGATATGATTTGGAAGAAAGGGCATCTCTCGTTATGCTTTTGC 1200
;
; QY 1201 GAAAGATTTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260
; DB 1201 GAAAGATTTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260
;
; QY 1261 TGTTTCAACACATTTAACTGCTTAAAGAAATGTTGCTCTCTGTCGACCAAGAAATTTTA 1320
; DB 1261 TGTTTCAACACATTTAACTGCTTAAAGAAATGTTGCTCTCTGTCGACCAAGAAATTTTA 1320
;
; QY 1321 TGAATAAGCATATAATCTGTTAACTGAGGCTCATTAAGGCTTGTGTTGNAATAAAGGCTG 1380
; DB 1321 TGAATAAGCATATAATCTGTTAACTGAGGCTCATTAAGGCTTGTGTTGNAATAAAGGCTG 1380
;
; QY 1381 TAAATCTGATTTTCAAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA 1440
; DB 1381 TAAATCTGATTTTCAAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA 1440
;
; QY 1441 TAAAGAAAATTTGGTAGATGATTTTATGGCAATCTTAGCAACCAATTTACCCATCCAGAGG 1500
; DB 1441 TAAAGAAAATTTGGTAGATGATTTTATGGCAATCTTAGCAACCAATTTACCCATCCAGAGG 1500
;
; QY 1501 ACTTGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAGGTTGTTGCTCAAT 1560
; DB 1501 ACTTGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAGGTTGTTGCTCAAT 1560
;
; QY 1561 AGCTGATAGTATACAACTGAGTGGTTATCTTTTGTGATGAACATGATATAATCAGTGA 1620

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 22:39:28 : Search time 170 Seconds
(without alignments)
7798.692 Million cell updates/sec

Title: US-09-765-271-55
Perfect score: 2389
Sequence: 1 TTCCTACAGTTGGACTGT.....TAAGTAAGGAAATAAATAAC 2389

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PTUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 2389 | 100.0 | 2389 | 3 | US-08-961-083-55 |
| 2 | 2389 | 100.0 | 2389 | 4 | US-09-536-784-55 |
| 3 | 2389 | 100.0 | 2451 | 4 | US-09-468-656A-9 |
| 4 | 2338 | 97.9 | 8195 | 4 | US-08-961-527-94 |
| 5 | 53 | 2.2 | 2290 | 3 | US-08-961-083-65 |
| 6 | 53 | 2.2 | 2290 | 4 | US-09-536-784-65 |
| 7 | 53 | 2.2 | 2359 | 4 | US-08-961-527-243 |
| 8 | 53 | 2.2 | 2531 | 4 | US-09-468-656A-5 |
| 9 | 53 | 2.2 | 2531 | 4 | US-09-468-656A-11 |
| 10 | 44 | 1.8 | 973 | 4 | US-08-961-527-355 |
| 11 | 30 | 1.3 | 40 | 3 | US-08-961-083-282 |
| 12 | 30 | 1.3 | 40 | 4 | US-09-536-784-282 |
| 13 | 30 | 1.3 | 40 | 4 | US-09-468-656A-3 |
| 14 | 27 | 1.1 | 36 | 4 | US-09-468-656A-1 |
| 15 | 27 | 1.1 | 37 | 3 | US-08-961-083-281 |
| 16 | 27 | 1.1 | 37 | 4 | US-09-536-784-281 |
| 17 | 26 | 1.1 | 35 | 4 | US-09-468-656A-2 |
| 18 | 24 | 1.0 | 1342 | 3 | US-08-961-083-181 |
| 19 | 24 | 1.0 | 1342 | 4 | US-09-536-784-181 |
| 20 | 24 | 1.0 | 1455 | 4 | US-09-468-656A-7 |
| 21 | 24 | 1.0 | 6867 | 4 | US-08-961-527-192 |
| 22 | 20 | 0.8 | 768 | 4 | US-09-489-039A-6492 |
| 23 | 19 | 0.8 | 622 | 3 | US-08-998-416-895 |
| 24 | 19 | 0.8 | 7493 | 1 | US-08-474-503-5 |
| 25 | 19 | 0.8 | 7493 | 1 | US-08-474-503-5 |
| 26 | 19 | 0.8 | 7493 | 2 | US-08-670-707A-5 |
| 27 | 19 | 0.8 | 7493 | 3 | US-09-037-601-5 |

ALIGNMENTS

RESULT 1
US-08-961-083-55
; Sequence 55, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:

| | | | | | |
|----|-----|---------|---|---------------------|-------------------|
| 19 | 0.8 | 7493 | 4 | US-09-315-179-5 | Sequence 5, Appli |
| 28 | 0.8 | 7493 | 5 | PCT-US94-13200-5 | Sequence 5, Appli |
| 29 | 0.8 | 51 | 4 | US-09-443-199C-979 | Sequence 979, App |
| 30 | 0.8 | 499 | 4 | US-09-641-638-52 | Sequence 52, Appl |
| 31 | 0.8 | 499 | 4 | US-09-641-638-572 | Sequence 572, App |
| 32 | 0.8 | 499 | 4 | US-09-641-638-572 | Sequence 290, Ap |
| 33 | 0.8 | 651 | 4 | US-09-134-000C-2990 | Sequence 390, App |
| 34 | 0.8 | 659 | 4 | US-09-221-017B-390 | Sequence 5116, Ap |
| 35 | 0.8 | 975 | 4 | US-09-489-039A-5116 | Sequence 2361, Ap |
| 36 | 0.8 | 987 | 4 | US-09-107-532A-1110 | Sequence 1110, Ap |
| 37 | 0.8 | 1161 | 4 | US-09-107-532A-1110 | Sequence 1943, Ap |
| 38 | 0.8 | 1744 | 4 | US-09-833-381-1943 | Sequence 1566, Ap |
| 39 | 0.8 | 2163 | 4 | US-09-543-681A-1566 | Sequence 1, Appli |
| 40 | 0.8 | 1664976 | 4 | US-08-916-421B-1 | Sequence 10, Appl |
| 41 | 0.7 | 20 | 2 | US-08-448-386A-10 | Sequence 10, Appl |
| 42 | 0.7 | 20 | 3 | US-08-816-426-10 | Sequence 10, Appl |
| 43 | 0.7 | 20 | 5 | PCT-US93-12161-10 | Sequence 56, Appl |
| 44 | 0.7 | 24 | 3 | US-08-981-462-56 | Sequence 15, Appl |
| 45 | 0.7 | 238 | 4 | US-09-491-356C-15 | Sequence 13, Appl |
| 46 | 0.7 | 253 | 4 | US-09-491-356C-13 | Sequence 14, Appl |
| 47 | 0.7 | 265 | 4 | US-09-491-356C-14 | Sequence 16, Appl |
| 48 | 0.7 | 265 | 4 | US-09-491-356C-16 | Sequence 17, Appl |
| 49 | 0.7 | 265 | 4 | US-09-491-356C-17 | Sequence 18, Appl |
| 50 | 0.7 | 265 | 4 | US-09-491-356C-18 | Sequence 2, Appli |
| 51 | 0.7 | 321 | 1 | US-08-175-388-2 | Sequence 2, Appli |
| 52 | 0.7 | 321 | 2 | US-08-779-620-2 | Sequence 2, Appli |
| 53 | 0.7 | 321 | 2 | US-08-818-722-2 | Sequence 149, App |
| 54 | 0.7 | 322 | 3 | US-08-961-083-149 | Sequence 149, App |
| 55 | 0.7 | 322 | 4 | US-09-536-784-149 | Sequence 9828, Ap |
| 56 | 0.7 | 463 | 4 | US-09-621-976-9828 | Sequence 9325, Ap |
| 57 | 0.7 | 464 | 4 | US-09-621-976-9325 | Sequence 216, App |
| 58 | 0.7 | 470 | 4 | US-09-056-556-216 | Sequence 211, App |
| 59 | 0.7 | 470 | 4 | US-09-072-596-211 | Sequence 216, App |
| 60 | 0.7 | 470 | 4 | US-09-072-967-216 | Sequence 1769, Ap |
| 61 | 0.7 | 524 | 4 | US-09-833-381-1769 | Sequence 1953, Ap |
| 62 | 0.7 | 633 | 4 | US-09-134-001C-1953 | Sequence 1007, Ap |
| 63 | 0.7 | 658 | 4 | US-09-221-017B-1007 | Sequence 1650, Ap |
| 64 | 0.7 | 687 | 4 | US-09-328-352-1650 | Sequence 1, Appli |
| 65 | 0.7 | 688 | 3 | US-08-981-462-1 | Sequence 60, Appl |
| 66 | 0.7 | 706 | 4 | US-09-470-191-60 | Sequence 373, App |
| 67 | 0.7 | 762 | 4 | US-09-071-035-373 | Sequence 2131, Ap |
| 68 | 0.7 | 789 | 4 | US-09-134-000C-2131 | Sequence 1170, Ap |
| 69 | 0.7 | 855 | 4 | US-09-134-000C-1170 | Sequence 1, Appli |
| 70 | 0.7 | 876 | 4 | US-09-501-558-1 | Sequence 3, Appli |
| 71 | 0.7 | 882 | 4 | US-09-501-558-3 | Sequence 1381, Ap |
| 72 | 0.7 | 909 | 4 | US-09-252-991A-1381 | Sequence 3547, Ap |
| 73 | 0.7 | 909 | 4 | US-09-328-352-3547 | Sequence 2698, Ap |
| 74 | 0.7 | 975 | 4 | US-09-107-532A-2698 | Sequence 6, Appli |
| 75 | 0.7 | 1077 | 3 | US-08-839-711-6 | Sequence 3376, Ap |
| 76 | 0.7 | 1080 | 4 | US-09-328-352-3376 | Sequence 1287, Ap |
| 77 | 0.7 | 1092 | 4 | US-09-252-991A-1287 | Sequence 1184, Ap |
| 78 | 0.7 | 1230 | 4 | US-09-252-991A-1184 | Sequence 223, App |
| 79 | 0.7 | 1311 | 4 | US-09-543-681A-223 | Sequence 9, Appli |
| 80 | 0.7 | 1543 | 4 | US-09-359-301A-9 | Sequence 2089, Ap |
| 81 | 0.7 | 1578 | 4 | US-09-107-532A-2089 | Sequence 3, Appli |
| 82 | 0.7 | 1623 | 4 | US-09-613-303-3 | Sequence 3, Appli |
| 83 | 0.7 | 1623 | 4 | US-10-267-311-3 | Sequence 3, Appli |
| 84 | 0.7 | 1654 | 2 | US-08-487-596-3 | Sequence 3, Appli |
| 85 | 0.7 | 1756 | 2 | US-08-466-589-3 | Sequence 3, Appli |
| 86 | 0.7 | 1756 | 2 | US-08-700-636-3 | Sequence 3, Appli |
| 87 | 0.7 | 1756 | 3 | US-08-467-574-3 | Sequence 3, Appli |
| 88 | 0.7 | 1756 | 3 | US-09-217-345-3 | Sequence 3, Appli |
| 89 | 0.7 | 1756 | 4 | US-09-892-985-3 | Sequence 807, App |
| 90 | 0.7 | 1857 | 4 | US-09-023-655-807 | |


```

source
1. .267
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0203J06"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/notes="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
Query Match 0.8%; Score 20; DB 28; Length 267;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1711 AGCCTACTAAGAAAAAG 1730
|||||
Db 71 AGCCTACTAAGAAAAAG 90

RESULT 25
BB803034 371 bp mRNA linear EST 19-NOV-2001
BB803034 RIKEN full-length enriched, 16 days neonate male
dienecephalon Mus musculus cDNA clone G630051N24 3', mRNA sequence.
BB803034
BB803034.1 GI:16975663
EST.
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 371)
REFERENCE
AUTHORS
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,I., Tomaru,A., Toya,T., Watanishi,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/

```

```

FEATURES
            source
            1. .371
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="G630051N24"
            /sex="male"
            /tissue_type="dienecephalon"
            /dev_stage="16 days neonate"
            /clone_lib="RIKEN full-length enriched, 16 days neonate
            male dienecephalon"

ORIGIN
Query Match 0.8%; Score 20; DB 10; Length 371;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 TTTCCTATATAGATGGA AAA 73
|||||
Db 200 TTTCCTATATAGATGGA AAA 219
|||||

Search completed: October 1, 2004, 03:33:05
Job time : 6095 secs

```

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

```

e mouse tissues.
            Location/Qualifiers
            1. .371
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="G630051N24"
            /sex="male"
            /tissue_type="dienecephalon"
            /dev_stage="16 days neonate"
            /clone_lib="RIKEN full-length enriched, 16 days neonate
            male dienecephalon"

```

```

source
1. .228
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0341023"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
Query Match 0.8%; Score 20; DB 28; Length 228;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 TTCTCTATATAGATGAAAA 73
Db 220 TTCTCTATATAGATGAAAA 201

RESULT 23
AZ769909
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

AZ769909 233 bp DNA linear GSS 16-FEB-2001
M0571J01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0571J01 F, genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

AZ769909.1 GI:12890543
GSS.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 233)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0571 row: J column: 01
Seq primer: CGTTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 233.
Location/Qualifiers

QY 318 TTATCAAGGTAGATGAAAA 337
Db 73 TTATCAAGGTAGATGAAAA 92

RESULT 24
AZ943045
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

AZ943045 267 bp DNA linear GSS 26-APR-2001
2M0203J06R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0203J06 R, genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

AZ943045.1 GI:13806817
GSS.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 267)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0203 row: J column: 06
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 267.
Location/Qualifiers

```

constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylnated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996) 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 935;
Best Local Similarity 100.0%; Pred.No. 72;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAGCAGAAGTTT 2210

db
500 AACTCAAAGCAGAGTTT 480

| RESULT 21 | LOCUS | 175 bp | mRNA | linear | EST 31-JUL-2000 |
|-----------|------------|--|------|--------|-----------------|
| B8538572 | B8538572 | RIKEN full-length enriched, 0 day neonate eyeball Mus musculus cDNA clone E13001212 3', mRNA sequence. | | | |
| | DEFINITION | | | | |

| | | |
|-----------|------------|------------|
| ACCESSION | BB538572 | |
| VERSION | BB538572.1 | GI:9594072 |

KEYWORDS EST. *Mus musculus* (house mouse)
SOURCE

| ORGANISM | Mus musculus |
|----------|--------------|
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| 16 | |
| 17 | |
| 18 | |
| 19 | |
| 20 | |
| 21 | |
| 22 | |
| 23 | |
| 24 | |
| 25 | |
| 26 | |
| 27 | |
| 28 | |
| 29 | |
| 30 | |
| 31 | |
| 32 | |
| 33 | |
| 34 | |
| 35 | |
| 36 | |
| 37 | |
| 38 | |
| 39 | |
| 40 | |
| 41 | |
| 42 | |
| 43 | |
| 44 | |
| 45 | |
| 46 | |
| 47 | |
| 48 | |
| 49 | |
| 50 | |
| 51 | |
| 52 | |
| 53 | |
| 54 | |
| 55 | |
| 56 | |
| 57 | |
| 58 | |
| 59 | |
| 60 | |
| 61 | |
| 62 | |
| 63 | |
| 64 | |
| 65 | |
| 66 | |
| 67 | |
| 68 | |
| 69 | |
| 70 | |
| 71 | |
| 72 | |
| 73 | |
| 74 | |
| 75 | |
| 76 | |
| 77 | |
| 78 | |
| 79 | |
| 80 | |
| 81 | |
| 82 | |
| 83 | |
| 84 | |
| 85 | |
| 86 | |
| 87 | |
| 88 | |
| 89 | |
| 90 | |
| 91 | |
| 92 | |
| 93 | |
| 94 | |
| 95 | |
| 96 | |
| 97 | |
| 98 | |
| 99 | |
| 100 | |

REFERENCE
1 (bases 1 to 175)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, I., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takanashi, P., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Havashizaki, Y.

```

ORGANISM      Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 799)
AUTHORS      Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE        A Comprehensive Collection of Chicken cDNAs
JOURNAL      Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE      22335534
PUBMED
COMMENT      Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES             Location/Qualifiers
     source
     1..799
     /organism="Gallus gallus"
     /mol_type="mRNA"
     /strain="Compton Line 151"
     /db_xref="taxon:9031"
     /clone="ChEST504c11"
     /sex="Female"
     /dev_stage="adult"
     /clone_lib="DH10B"
     /note="Organ: heart; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN
Query Match      0.9%; Score 21; DB 13; Length 799;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2190 AACTCAAGACGACGAGTTT 2210
      |||||||||||||||||||
DB      465 AACTCAAGACGACGAGTTT 445

RESULT 19
CC088871
LOCUS      CC088871
DEFINITION      CSU-K33r.8E23.SP6 CSU-K33r Aedes aegypti genomic clone
                CSU-K33r.8E23, genomic survey sequence.
ACCESSION      CC088871
VERSION
KEYWORDS
SOURCE
ORGANISM      Aedes aegypti (yellow fever mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
Stegomyia.
REFERENCE
1 (bases 1 to 870)
AUTHORS      Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
TITLE        End sequencing of Aedes aegypti BACs
JOURNAL      Unpublished (2003)
COMMENT      Other_GSSs: CSU-K33r.8E23.T7

ORGANISM      Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 935)
AUTHORS      Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE        A Comprehensive Collection of Chicken cDNAs
JOURNAL      Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE      22335534
PUBMED
COMMENT      Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES             Location/Qualifiers
     source
     1..935
     /organism="Gallus gallus"
     /mol_type="mRNA"
     /strain="Compton Line 151"
     /db_xref="taxon:9031"
     /clone="ChEST486f8"
     /sex="Female"
     /tissue_type="cerebrum"
     /dev_stage="adult"
     /lab_host="DH10B"
     /clone_lib="CSQCCHN53"
     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was

```

```

Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado
State University.
Seq primer: SP6
Class: BAC ends.

FEATURES             Location/Qualifiers
     source
     1..870
     /organism="Aedes aegypti"
     /mol_type="genomic DNA"
     /strain="Rexville"
     /db_xref="taxon:7159"
     /clone="CSU-K33r.8E23"
     /clone_lib="CSU-K33r"
     /note="Vector: pBelOBAC11; Site_1: HindIII"

ORIGIN
Query Match      0.9%; Score 21; DB 28; Length 870;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1222 TGTAAAAAATCTTGAAGCAA 1242
      |||||||||||||||||||
DB      91 TGTAAAAAATCTTGAAGCAA 111

RESULT 20
BU273390/c
LOCUS      BU273390
DEFINITION      603531527F1 CSEQCHN53 Gallus gallus cDNA clone ChEST486f8 5', mRNA
                sequence.
ACCESSION      BU273390
VERSION
KEYWORDS
SOURCE
ORGANISM      Gallus gallus (chicken)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 935)
AUTHORS      Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE        A Comprehensive Collection of Chicken cDNAs
JOURNAL      Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE      22335534
PUBMED
COMMENT      Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES             Location/Qualifiers
     source
     1..935
     /organism="Gallus gallus"
     /mol_type="mRNA"
     /strain="Compton Line 151"
     /db_xref="taxon:9031"
     /clone="ChEST486f8"
     /sex="Female"
     /tissue_type="cerebrum"
     /dev_stage="adult"
     /lab_host="DH10B"
     /clone_lib="CSQCCHN53"
     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was

```


COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@gene.nig.ac.jp.

FEATURES source
1. .690
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Haruna Nijo"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="bags35c04"
/tissue_type="shoots"
/dev_stage="germination"
/clone_lib="K. Sato unpublished cDNA library, cv. Haruna
Nijo germination shoots"

ORIGIN

Query Match 0.9%; Score 21; DB 12; Length 690;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1078 AGGCCGCAACTGCACCAAA 1098
|||||
DB 76 AGGCCGCAACTGCACCAAA 96

RESULT 14

BU403931/c
LOCUS BU403931 731 bp mRNA linear EST 27-NOV-2002
DEFINITION 604141302F1 CSEQCHN59 Gallus gallus cDNA clone CHEST974d2 5', mRNA
sequence.

ACCESSION BU403931
VERSION BU403931.1 GI:25772987
KEYWORDS EST.
SOURCE Gallus gallus (chicken)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE

1 (bases 1 to 731)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392

COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)

PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source
1. .731
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST974d2"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHN59"

/note="Organ: limbs; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction."

Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 731;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGACGACGAAGTTT 2210
|||||
DB 531 AACTCAAGACGACGAAGTTT 511

RESULT 15

BU283494/c
LOCUS BU283494 742 bp mRNA linear EST 27-NOV-2002
DEFINITION 603862178F1 CSEQCHN54 Gallus gallus cDNA clone CHEST875k20 5', mRNA
sequence.

ACCESSION BU283494
VERSION BU283494.1 GI:25732950
KEYWORDS EST.
SOURCE Gallus gallus (chicken)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE

1 (bases 1 to 742)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392

COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)

PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source
1. .742
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST875k20"
/sex="Female"
/tissue_type="not cerebrum or cerebellum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN54"

/note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction."

Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer

Query Match 0.9%; Score 21; DB 28; Length 640;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1539 ACGAAGTTCGATTGCTCAAT 1559
 |||||
 Db 332 ACGAAGTTCGATTGCTCAAT 312

RESULT 11
 BU345704/c
 LOCUS
 DEFINITION 604170839F1 CSEQCHN68 Gallus gallus CDNA clone CHES1033m10 5',
 mRNA sequence.
 ACCESSION BU345704
 VERSION BU345704.1 GI:25853705
 KEYWORDS EST.
 SOURCE
 ORGANISM Gallus gallus (chicken)
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 664)
 AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..664
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="ChEST1033m10"
 /sex="Female"
 /tissue_type="cerebellum"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN68"
 /notes="Organ: brain; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN
 Query Match 0.9%; Score 21; DB 13; Length 664;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAGCAGAGATT 2210
 |||||
 Db 295 AACTCAAGAGCAGAGATT 275

RESULT 12
 AG084591/c
 LOCUS
 DEFINITION Pan troglodytes DNA, clone: PTB-082G04.F, genomic survey sequence.
 ACCESSION AG084591
 VERSION AG084591.1 GI:16636393
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes

REFERENCE 1
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE BAC end sequences of Library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 681)
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 [E-mail:chimpes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170]
 COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.

PRIMERS

LIBRARY
 Sequencing: -21M13
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.

FEATURES
 source
 1..681
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="PTB-082G04.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
 Query Match 0.9%; Score 21; DB 29; Length 681;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 CAGATGTTAAGCAATCCAA 1775
 |||||
 Db 142 CAGATGTTAAGCAATCCAA 122

RESULT 13
 BU549886
 LOCUS
 DEFINITION BU549886 K. Sato unpublished cDNA library, cv. Haruna Niho
 germination shoots Hordeum vulgare subsp. vulgare cDNA clone
 bags35c04 3', mRNA sequence.
 ACCESSION BU549886
 VERSION BU549886.1 GI:24968336
 KEYWORDS EST.
 SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare

REFERENCE 1
 AUTHORS Sato,K., Saisho,D. and Takeda,K.
 TITLE Barley EST sequencing project in NIG and Okayama Univ
 JOURNAL Unpublished (2002)

/db_xref="taxon:7159"
 /clone="Notre Dame Liverpool-9714"
 /clone_lib="Notre Dame Liverpool"
 /note="Vector: pECBAC1; Site_1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"

ORIGIN

Query Match 0.9%; Score 21; DB 29; Length 551;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1222 TGTAAATAATCTGAAGCAA 1242
 DB 27 TGTAAATAATCTGAAGCAA 7

RESULT 9

BU397983/c
 LOCUS BU397983 581 bp mRNA linear EST 27-NOV-2002
 DEFINITION 603534550F1 CSEQCHN58 Gallus gallus cDNA clone ChEST494c13 5', mRNA sequence.

ACCESSION BU397983
 VERSION BU397983
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 581)
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392

COMMENT

Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)

PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source

Location/Qualifiers
 1..581
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="ChEST494c13"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN58"

/note="Organ: small intestine; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 581;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGACGACGAAGTTT 2210
 DB 368 AACTCAAGACGACGAAGTTT 348

RESULT 10

AZ391393/c
 LOCUS AZ391393 640 bp DNA linear GSS 03-OCT-2000
 DEFINITION 1M0153A03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0153A03 R, genomic survey sequence.

ACCESSION AZ391393
 VERSION AZ391393
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 640)

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL COMMENT

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0153 row: A column: 03

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 640.

FEATURES

source

Location/Qualifiers
 1..640
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0153A03"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: FWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN


```

RESULT 6
BI544663          499 bp      mRNA      linear      EST 05-SEP-2001
LOCUS             603242619F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5284997 5',
DEFINITION        mRNA sequence.
ACCESSION         BI544663
VERSION           BI544663.1 GI:15431975
KEYWORDS          EST.
SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE         1 (bases 1 to 499)
AUTHORS           NIH-MGC http://mgc.nci.nih.gov/.
TITLE            National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL           Unpublished (1999)
COMMENT          Contact: Robert Strausberg, Ph.D.
                  Email: cgapbs-r@mail.nih.gov
                  Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
                  cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
                  Toshiyuki and Piero Carninci (RIKEN)
                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLNL1720 row: f column: 06
                  High quality sequence stop: 299.
FEATURES          Location/Qualifiers
                    1..499
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5284997"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /notes="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTNN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."
ORIGIN
Query Match      0.9%; Score 21; DB 12; Length 499;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 389 GAGGAATCAATCGACAAAAA 409
      |||||
Db 414 GAGGAATCAATCGACAAAAA 434

RESULT 7
AW932212
LOCUS             AW932212
DEFINITION        EST358055 tomato fruit mature green, TAMU Lycopersicon esculentum
                  cDNA clone cLEF47L13 5', mRNA sequence.
ACCESSION         AW932212
VERSION           AW932212.1 GI:8107613
KEYWORDS          EST.
SOURCE            Lycopersicon esculentum (tomato)
ORGANISM          Lycopersicon esculentum
                  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE         1 (bases 1 to 549)

```

```

AUTHORS           Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Liang, F.,
                  Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ann, S.,
                  Ronning, C.M., Fraser, C.M., Martin, G.B., Tankley, S.D. and
                  Giovannoni, J.
TITLE            Generation of ESTs from tomato fruit tissue
JOURNAL           Unpublished (1999)
COMMENT          Contact: CUGI
                  Clemson University Genomics Institute
                  Clemson University
                  100 Jordan Hall, Clemson, SC 29634, USA
                  Email: http://www.genome.clemson.edu/orders/index.html
                  5 prime sequence.
FEATURES          Location/Qualifiers
                    1..549
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEF47L13"
                     /tissue_type="fruit pericarp"
                     /dev_stage="mature green (3-5 days pre-ripening)"
                     /lab_host="SOLR"
                     /clone_lib="tomato fruit mature green, TAMU"
                     /notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; cLEF- Fruit were tagged at the 1cm stage and
                     harvested 3-5 days prior to ripening. Fruit were cut in
                     half to verify the seeds were indeed 'immature' and the
                     seeds and locules were discarded prior to freezing the
                     pericarp"
ORIGIN
Query Match      0.9%; Score 21; DB 10; Length 549;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 AGGTGGATATGTTATCAAGG 326
      |||||
Db 135 AGGTGGATATGTTATCAAGG 155

RESULT 8
CC847187
LOCUS             CC847187/c
DEFINITION        ND.L.9714.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
                  Notredame Liverpool-9714, genomic survey sequence.
ACCESSION         CC847187
VERSION           CC847187.1 GI:33194851
KEYWORDS          GSS.
SOURCE            Aedes aegypti (yellow fever mosquito)
ORGANISM          Aedes aegypti
                  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                  Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
                  Stegomyia.
REFERENCE         1 (bases 1 to 551)
AUTHORS           Loftus, B., Shetty, J., Knudson, D. and Severson, D.
TITLE            BAC end sequencing of Aedes aegypti
JOURNAL           Unpublished (2003)
COMMENT          Other_GSSs: NDL.9714.T7
                  Contact: Brendan Loftus
                  Department of Eukaryotic Genomics
                  TIGR
                  9712 Medical Center Drive, Rockville, MD 20850, USA
                  Tel: 301-838-3543
                  Fax: 301-838-0208
                  Email: enta@tigr.org
                  Library was provided by David Severson
                  Seq primer: SP6
                  Class: BAC ends.
FEATURES          Location/Qualifiers
                    1..551
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="liverpool"

```

3-1-1 Koyadai, Teukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda PLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

Location/Qualifiers
1..618
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAF15-41-A11"
/tissue_type="mixture of silique and flower"
/lab_host="DH10B"
/clone_lib="RAF15"
/note="Site_1: BamHI; Site_2: SalI"

ORIGIN

Query Match 0.9%; Score 22; DB 9; Length 618;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2295 ACTCTTCAAAATTCGATAACA 2306

DB 481 ACTCTTCAAAATTCGATAACA 502

RESULT 4

BH970231/c
LOCUS odf81d11.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION sequence.
ACCESSION BH970231
VERSION BH970231.1 GI:23453048
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 682)
Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash,W., Rabinowicz,P.D. and Wilson,R.K.
Whole genome shotgun reads from Brassica oleracea
Unpublished (2002)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: odf81 row: d column: 11
Seq primer: -21UpPOT forward
Class: shotgun
High quality sequence start: 14
High quality sequence stop: 519.
Location/Qualifiers
1..682
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pORW13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea TO1000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

FEATURES

Location/Qualifiers
1..682
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pORW13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea TO1000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 411;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAAGCAGAGTTT 2210

DB 273 AACTCAAGAAGCAGAGTTT 253

Query Match 0.9%; Score 22; DB 28; Length 682;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1811 AAAGGGGAAAACGAATTCAC 1832

DB 44 AAAGGGGAAAACGAATTCAC 23

RESULT 5

BU473013/c
LOCUS 603365019F1 CSEQRBN21 Gallus gallus linear EST 30-NOV-2002
DEFINITION sequence.
ACCESSION BU473013
VERSION BU473013.1 GI:25966590
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 411)
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445332
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

Location/Qualifiers
1..411
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer"
/db_xref="taxon:9031"
/clone="ChEST262f5"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQRBN21"
/note="Organ: ovary; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 411;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAAGCAGAGTTT 2210

DB 273 AACTCAAGAAGCAGAGTTT 253

78 20 0.8 923 10 BG033498
 79 20 0.8 924 29 CG862220
 80 20 0.8 924 29 CG862220
 81 20 0.8 957 13 BQ712408
 82 20 0.8 984 29 CG858044
 83 20 0.8 2185 11 AK037932
 84 19 0.8 66 28 BH907895
 85 19 0.8 110 28 CC021347
 86 19 0.8 136 14 H74951
 87 19 0.8 158 29 CE567072
 88 19 0.8 175 10 BE505229
 89 19 0.8 186 28 BH867080
 90 19 0.8 220 9 AV236691
 19 0.8 250 14 CF765303

ALIGNMENTS

RESULT 1
 CE751062/c
 LOCUS
 DEFINITION tigr-gss-dog-17000369611786 Dog Library Canis familiaris genomic,
 genomic survey sequence.
 CE751062
 CE751062.1 GI:37091679
 GSS.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 REFERENCE 1 (bases 1 to 469)
 AUTHORS Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
 Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
 Venter, J.C.
 TITLE The dog genome: survey sequencing and comparative analysis
 JOURNAL Science 301 (5641), 1898-1903 (2003)
 MEDLINE 22875432
 PUBMED 14512627
 COMMENT Contact: Kirkness EF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirknes@tigr.org
 Class: shotgun.

FEATURES
 source
 1..469
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /strain="Standard Poodle"
 /db_xref="taxon:9615"
 /clone_lib="Dog Library"
 /note="Site 1: BstXI; Libraries were prepared from
 peripheral blood"

ORIGIN
 Query Match 1.0%; Score 23; DB 29; Length 469;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2140 ACCTGCTGAGCCAGAGTCCCTC 2162
 |||||
 Db 269 ACCTGCTGAGCCAGAGTCCCTC 247

RESULT 2
 BE528097
 LOCUS
 DEFINITION M70N06STM Arabidopsis developing seed Arabidopsis thaliana cDNA
 clone 600037123R1 5', mRNA sequence.
 ACCESSION BE528097
 VERSION BE528097.1 GI:9786075

KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE 1 (bases 1 to 261)
 AUTHORS White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
 Illardua, O., Jaworski, J.G., Ohlgrogge, J., and Benning, C.
 TITLE A new set of Arabidopsis expressed sequence tags from developing
 seeds. The metabolic pathway from carbohydrates to seed oil
 JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
 MEDLINE 20567808
 PUBMED 11115876
 COMMENT Contact: Benning, C
 Dept. of Biochemistry & Molecular Biology
 Michigan State University
 224 Biochemistry, Michigan State University, East Lansing, MI
 48824, USA
 Tel: 517 355 1609
 Fax: 517 353 9334
 Email: benning@msu.edu
 Clones were originally prepared at Michigan State University.
 Arabidopsis Biological Resource Center, The Ohio State University,
 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
 USA, FAX: 6142920603 TEL: 6142929371.

FEATURES
source

1..261
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="600037123R1"
 /tissue_type="seed"
 /dev_stage="5-13 days after flowering"
 /lab_host="E.coli"
 /clone_lib="Arabidopsis developing seed"
 /note="Organ: Developing seed; Vector: pBluescript SK-;
 Site_1: EcoRI; Site_2: XhoII"

ORIGIN

Query Match 0.9%; Score 22; DB 10; Length 261;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2285 ACTCTCAAAATTATGATAACA 2306
 |||||
 Db 200 ACTCTCAAAATTATGATAACA 221

RESULT 3
AU236995

LOCUS
 DEFINITION AU236995 RAFL15 Arabidopsis thaliana cDNA clone RAFL15-41-A11 5',
 mRNA sequence.

ACCESSION AU236995
 VERSION AU236995.1 GI:19876164

KEYWORDS
SOURCE
ORGANISM

Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS

1 (bases 1 to 618)
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
 Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, J.,
 Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A.,
 Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
 TITLE Large scale analysis of Arabidopsis full-length cDNA
 JOURNAL Unpublished (2002)
 COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 22:20:23 : Search time 6083 Seconds
(without alignments)
11727.889 Million cell updates/sec

Title: US-09-765-271-55
Perfect score: 2389
Sequence: 1 TTCTTACGAGTGGACTGT.....TAAGTAAGGAAAAATAAAC 2389

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 27513289 seqs, 14931090276 residues
Word size : 0
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| C 1 | 23 | 1.0 | 459 | 29 | CE751062 tigr-gss- |
| C 2 | 22 | 0.9 | 261 | 10 | BE528097 M70N06STM |
| C 3 | 22 | 0.9 | 618 | 9 | AU236995 |
| C 4 | 22 | 0.9 | 682 | 28 | BH970231 odff81d11. |


```
DT 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae BVH-11-2 gene SEQ ID NO:13.
DE
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; diagnosis; meningitis; bacteremia;
KW otitis media; pneumonia; immunisation; bactericidal; ds.
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
PN
XX 06-JUL-2000.
PD
XX 20-DEC-1999; 99WO-CA001218.
PF
XX 23-DEC-1998; 98US-0113800P.
PR
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI
XX WPI; 2000-452397/39.
DR
XX P-PSDB; AAB12720.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteremia and/or pneumonia.
XX
XX Example 6; Fig 16; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteremia and/or
CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11-2
CC protein antigen
XX
XX Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;
SQ
Query Match 2.2%; Score 53; DB 3; Length 2639;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTTGGTACGAGATTCAAG 1035
Db 1113 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTTGGTACGAGATTCAAG 1165

RESULT 25
ABK15104
ID ABK15104 standard; DNA; 2639 BP.
XX
XX ABK15104;
XX
XX 08-MAY-2002 (first entry)
DT
XX
XX DNA encoding Streptococcus pneumoniae BVH-11-2.
DE
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteremia; pneumonia;
KW streptococcal bacterial infection; gene; ds; BVH-11-2.
XX
XX Streptococcus pneumoniae.
OS
XX
XX Key Location/Qualifiers
FH 114..2630
FT CDS
FT /tag= a
FT /product= "BVH-11-2"
FT /note= "The gene is flanked by sequences from the vector
FT SP64, no information on which is given in the
FT specification"
```

```
XX WO200198334-A2.
PN
XX 27-DEC-2001.
PD
XX 19-JUN-2001; 2001WO-CA000908.
PF
XX 20-JUN-2000; 2000US-0212683P.
PR
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI
XX WPI; 2002-122272/16.
DR
XX P-PSDB; AAU75934.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
PT polypeptides, useful as vaccine components for treating or preventing
PT streptococcal infections such as otitis media, meningitis, and
PT bacteremia.
XX
XX Example 3; Fig 5; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90% identity
CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteremia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection. A
CC polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence encodes the antigenic
CC Streptococcus pneumoniae protein BVH-11-2, used to create the antigenic
CC peptides described in the method of the invention
XX
XX Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;
SQ
Query Match 2.2%; Score 53; DB 6; Length 2639;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTTGGTACGAGATTCAAG 1035
Db 1113 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTTGGTACGAGATTCAAG 1165

Search completed: September 30, 2004, 23:15:41
Job time : 923 secs
```

```
Db      1003 GCTCGTATTATTCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1055
|||||
RESULT 22
AAAA7605
ID      AAAA7605 standard; DNA; 2531 BP.
XX
AC      AAAA7605;
XX
XX      20-OCT-2000 (first entry)
XX
DE      Recombinant variant of Sp36 gene (Sp36B) of S. pneumoniae.
XX
KW      Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW      histidine triad residue; Sp36; antibody; otitis media;
KW      nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW      meningitis; lobar pneumonia; ds.
XX
OS      Streptococcus pneumoniae.
XX
XX      Key      Location/Qualifiers
FH      CDS      1..2531
FT      /*tag= a
FT      /product= "Sp36B polypeptide"
XX
XX      WO200037105-A2.
XX
XX      29-JUN-2000.
XX
XX      21-DEC-1999; 99WO-US030390.
XX
XX      21-DEC-1998; 98US-0113048P.
XX
XX      (MEDI-) MEDIMMUNE INC.
XX
XX      Johnson LS, Koenig S, Adamou JE;
XX
XX      WPI; 2000-452129/39.
XX      P-PSDB; AAB01469.
XX
XX      Vaccine useful for prophylaxis and treatment of pneumococcal infections
XX      such as otitis media, nasopharyngeal and bronchial infections, comprises
XX      Streptococcus pneumoniae proteins.
XX
XX      Disclosure; Page 69-70; 70pp; English.
XX
XX      Although a number of proteins have been suggested as being involved in
XX      the pathogenicity of Streptococcus pneumoniae, there still remains a need
XX      to identify polypeptides having epitopes in common from various strains
XX      of S. pneumoniae in order to utilize such polypeptides in vaccines to
XX      protect against a wide variety of S. pneumoniae. New vaccine compositions
XX      are described which comprise a Streptococcus pneumoniae polypeptide (or
XX      fragments) of 80 - 680 amino acids in length that comprise at least one
XX      histidine triad residue (HxxHxxH) or a coiled-coil region, or an antibody
XX      directed against these features. The vaccine is useful in protecting
XX      against infection by Streptococcus pneumoniae. The vaccine composition
XX      comprising antibodies to is useful for passive immunization for treating
XX      Pneumococcal infections which includes otitis media, nasopharyngeal and
XX      bronchial infections
XX
XX      Sequence 2531 BP; 861 A; 508 C; 550 G; 609 T; 0 U; 3 Other;
XX
XX      Query Match      2.2%; Score 53; DB 3; Length 2531;
XX      Best Local Similarity 100.0%; Pred. No. 6.8e-16;
XX      Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY      983 GCTCGTATTATTCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1035
XX      |||||||
XX      Db      1000 GCTCGTATTATTCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1052
XX      |||||||
XX
XX      RESULT 23
```

```
AAAA7602
ID      AAAA7602 standard; DNA; 2531 BP.
XX
AC      AAAA7602;
XX
XX      20-OCT-2000 (first entry)
XX
DE      Recombinant variant of Sp36 gene (Sp36D) of S. pneumoniae.
XX
KW      Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW      histidine triad residue; Sp36; antibody; otitis media;
KW      nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW      meningitis; lobar pneumonia; ds.
XX
OS      Streptococcus pneumoniae.
XX
XX      Key      Location/Qualifiers
FH      CDS      1..2517
FT      /*tag= a
FT      /product= "Sp36D polypeptide"
XX
XX      WO200037105-A2.
XX
XX      29-JUN-2000.
XX
XX      21-DEC-1999; 99WO-US030390.
XX
XX      21-DEC-1998; 98US-0113048P.
XX
XX      (MEDI-) MEDIMMUNE INC.
XX
XX      Johnson LS, Koenig S, Adamou JE;
XX
XX      WPI; 2000-452129/39.
XX      P-PSDB; AAB01466.
XX
XX      Vaccine useful for prophylaxis and treatment of pneumococcal infections
XX      such as otitis media, nasopharyngeal and bronchial infections, comprises
XX      Streptococcus pneumoniae proteins.
XX
XX      Disclosure; Page 57-58; 70pp; English.
XX
XX      Although a number of proteins have been suggested as being involved in
XX      the pathogenicity of Streptococcus pneumoniae, there still remains a need
XX      to identify polypeptides having epitopes in common from various strains
XX      of S. pneumoniae in order to utilize such polypeptides in vaccines to
XX      protect against a wide variety of S. pneumoniae. New vaccine compositions
XX      are described which comprise a Streptococcus pneumoniae polypeptide (or
XX      fragments) of 80 - 680 amino acids in length that comprise at least one
XX      histidine triad residue (HxxHxxH) or a coiled-coil region, or an antibody
XX      directed against these features. The vaccine is useful in protecting
XX      against infection by Streptococcus pneumoniae. The vaccine composition
XX      comprising antibodies to is useful for passive immunization for treating
XX      Pneumococcal infections which includes otitis media, nasopharyngeal and
XX      bronchial infections
XX
XX      Sequence 2531 BP; 836 A; 511 C; 547 G; 637 T; 0 U; 0 Other;
XX
XX      Query Match      2.2%; Score 53; DB 3; Length 2531;
XX      Best Local Similarity 100.0%; Pred. No. 6.8e-16;
XX      Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY      983 GCTCGTATTATTCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1035
XX      |||||||
XX      Db      1003 GCTCGTATTATTCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1055
XX      |||||||
XX
XX      RESULT 24
XX      AAA65737
XX      ID      AAA65737 standard; DNA; 2639 BP.
XX
XX      AC      AAA65737;
XX
XX
```

Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATTCGGTACCAGATTCAAG 1035
 |||||
 Db 1003 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATTCGGTACCAGATTCAAG 1055

RESULT 20
 AAA05417
 ID AAA05417 standard; DNA; 2481 BP.
 XX
 AC AAA05417;
 XX
 DT 24-MAY-2000 (first entry)
 XX
 DE Streptococcus pneumoniae nucleotide sequence ID311.
 XX
 KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200006737-A2.
 XX
 PD 10-FEB-2000.
 XX
 PF 27-JUL-1999; 99WO-GB002451.
 XX
 PR 27-JUL-1998; 98GB-00016337.
 FR 19-MAR-1999; 99US-0125164P.
 XX
 PA (MICR-) MICROBIAL TECHNIQS LTD.
 XX
 PI Gilbert CFG, Hansbro PW;
 XX
 DR WPI; 2000-195300/17.
 XX
 P-PSDB; AAY81662.

New Streptococcal protein, useful as a vaccine, for diagnosis of pneumococcal diseases and for screening agents capable of antagonizing or inhibiting expression of the protein.
 XX
 PS Claim 2; Page 99; 108pp; English.
 XX
 CC AAY81501 to AAY81679 represent specifically claimed protein sequences isolated from Streptococcus pneumoniae. AAY05407 to AAY05590 represent specifically claimed nucleotide sequences isolated from S. pneumoniae. The sequences have antibacterial and antiinflammatory properties. The protein sequences, and fragments of them, are useful as immunogens and in antigens. The nucleotide sequences can be used in vaccines and in diagnostic assays. The proteins and nucleotides can be useful for the detection and diagnosis of S. pneumoniae. The protein sequences are also useful for screening an agent capable of antagonising, inhibiting or interfering with the function or expression of the proteins in which the agent is useful for treatment or prophylaxis of S. pneumoniae infection and meningitis. AAY05591 to AAY05614 represent primers used in the exemplification of the present invention
 XX
 SQ Sequence 2481 BP; 839 A; 514 C; 538 G; 590 T; 0 U; 0 Other;

Query Match 2.2%; Score 53; DB 3; Length 2481;
 Best Local Similarity 100.0%; Pred. No. 6.8e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATTCGGTACCAGATTCAAG 1035
 |||||
 Db 1003 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATTCGGTACCAGATTCAAG 1055

RESULT 21
 ABX06705
 ID ABX06705 standard; DNA; 2517 BP.
 XX
 AC ABX06705;

XX 27-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain coding region #993.
 XX
 KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae; type 4 strain.
 XX
 PN WO200277021-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-IB002163.
 XX
 PR 27-MAR-2001; 2001GB-00007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masignani V, Tettelin H, Fraser C;
 XX
 DR WPI; 2003-040579/03.
 DR P-PSDB; ABU01418.
 XX
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX
 PS Claim 6; SEQ ID NO 1985; 56pp; English.

XX The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the Streptococcus pneumoniae type 4 strain genomic sequence appearing as ABS56434. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a Streptococcus nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a Streptococcus pneumoniae bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2489 identified coding region from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to standardise OS field)

SQ Sequence 2517 BP; 830 A; 509 C; 545 G; 633 T; 0 U; 0 Other;
 Query Match 2.2%; Score 53; DB 7; Length 2517;
 Best Local Similarity 100.0%; Pred. No. 6.8e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATTCGGTACCAGATTCAAG 1035

XX AB084792 to AB084904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
CC pneumoniae antigens have antibacterial activity and can be used in
CC vaccines. The S. pneumoniae antigens can also be used to prevent or
CC attenuate a Streptococcal infection in an animal. The polynucleotides
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
CC nucleic acids. AB084905 to AB085130 represent primers used in the cloning
CC of S. pneumoniae ORFs (open reading frames) which are used in an example
CC from the present invention
XX
SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;

Query Match 2.2%; Score 53; DB 6; Length 2290;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 983 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
|||
Db 944 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 996
|||

RESULT 16
ID ADC45146 standard; DNA; 2290 BP.
XX
AC ADC45146;
XX

DT 18-DEC-2003 (first entry)

DE S. pneumoniae DNA encoding antigen SP042.

KW Antigen; ds; bacterial infection; vaccine; pneumonia; antibacterial.

OS Streptococcus pneumoniae.

XX US6573082-B1.

XX 03-JUN-2003.

XX 28-MAR-2000; 2000US-00536784.

XX 31-OCT-1996; 96US-0029960P.

XX 30-OCT-1997; 97US-00961083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;

XX WPI; 2003-764574/72.

XX P-PSDB; ADC45147.

XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
PT useful for producing vaccines for prevention or attenuation of infection
PT by Streptococcus pneumoniae.

XX Example 1; SEQ ID NO 65; 58pp; English.

XX The invention relates to an isolated polynucleotide consisting of a
CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
CC antigens. Also included are making a recombinant vector by inserting the
CC nucleic acid into a vector, an isolated polynucleotide consisting of at
CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
CC acids are useful as DNA vaccine against Streptococcus pneumoniae
CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
CC antigen nucleic acids are useful as probes for use in diagnostic methods
CC for detecting S. pneumoniae gene expression. The present sequence encodes
CC an S. pneumoniae antigenic protein.

XX Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;

Query Match 2.2%; Score 53; DB 9; Length 2290;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 983 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
|||
Db 944 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 996
|||

RESULT 17
AAV52376
ID AAV52376 standard; DNA; 2359 BP.

XX AC AAV52376;

XX DT 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:243.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

XX WO9818931-A2.

XX PD 07-MAY-1998.

XX PF 30-OCT-1997; 97WO-US019588.

XX PR 31-OCT-1996; 96US-0029960P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
PI Dougherty BA;

XX WPI; 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.

XX Claim 1; Page 1265-1266; 1409pp; English.

XX The present invention describes a computer readable medium which has the
CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
CC on it, or a representative fragment or a sequence at least 9% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
CC 391, identifying members of the library which contain sequences that
CC hybridize to the target sequence and isolating the nucleic acid molecules
CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an
CC organism, amplifying nucleic acid molecules whose nucleotide sequence is
CC homologous to amplification primers derived from the fragment of the S.
CC pneumoniae genome to prime the amplification and isolating the amplified
CC sequences. The computer readable medium can be used in a computer-based
CC system for identifying fragments of the S. pneumoniae genome of
CC commercial importance, or expression modulating fragments of the S.
CC pneumoniae genome. Products from the present invention can be used in
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines
CC for S. pneumoniae

XX Sequence 2359 BP; 786 A; 451 C; 511 G; 611 T; 0 U; 0 Other;

Query Match 2.2%; Score 53; DB 2; Length 2359;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;

PS Claim 54; Page 52-54; 66pp; English.

XX This DNA sequence encodes a 79 kDa protein (see AAY05753) of

CC Streptococcus pneumoniae serotype 4 that is capable of degrading human

CC complement protein C3 (HCP3). It was identified in the S. pneumoniae

CC serotype 4 genome by alignment to another novel open reading frame (see

CC AAX25393) that codes for a 20 kDa HCP3 protease (AAY05752) of S.

CC pneumoniae serotype 3. This suggested the open reading frame that codes

CC for the 20 kDa protein may be part of a larger open reading frame. Amino

CC acids 1-58 and 90-132 of the 20 kDa protein have substantial sequence

CC identity with amino acids 170-227 and 258-300 of the 79 kDa protein.

CC Proteins and peptides or polypeptides containing these regions, and DNA

CC sequences encoding them (nucleotides 507-681 and 827-999 of the present

CC sequence) are claimed. HCP3 proteases and polypeptides can be used as

CC immune system stimulating compositions (claimed). They can produce an

CC immune response against S. pneumoniae to immunize or treat a mammalian

CC subject against infection or colonization (claimed). They can produce a B

CC cell response, a T cell response, an epithelial cell response, or an

CC endothelial cell response (claimed). The expression of the proteins on

CC the surface of an organ of an animal used in xenotransplantation can be

CC used to inhibit C3-mediated inflammation and rejection

XX

SQ Sequence 2163 BP; 737 A; 450 C; 462 G; 514 T; 0 U; 0 Other;

Query Match 2.2%; Score 53; DB 2; Length 2163;

Best Local Similarity 100.0%; Pred. No. 6.8e-16;

Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATTTGGTACCGATTCAAG 1035

DB 688 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATTTGGTACCGATTCAAG 740

RESULT 14

AAV27356

ID AAV27356 standard; DNA; 2290 BP.

XX

AC AAV27356;

XX

DT 02-OCT-1998 (first entry)

XX

DE Streptococcus pneumoniae SP0042 nucleotide.

XX

KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;

KW detection; pneumonia; otitis media; meningitis; ss.

XX

OS Streptococcus pneumoniae.

XX

FH Key Location/Qualifiers

FT CDS 2..2290

FT /*tag= a

FT /product= "SP0042"

FT /transl_except= (pos:152..154,aa:Xaa)

FT /transl_except= (pos:1406..1408,aa:Xaa)

FT /transl_except= (pos:1430..1432,aa:Xaa)

FT /note= "no stop codon given; Xaa is unspecified"

XX

PN WO9818930-A2.

XX

XX

PD 07-MAY-1998.

XX

XX

PF 30-OCT-1997; 97WO-US019422.

XX

XX

PR 31-OCT-1996; 96US-0029960P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;

XX

XX

XX WPI; 1998-272224/24.

DR F-PSDB; AAW55095.

XX

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae

PT - or their epitope-containing fragments, useful in protective or

PT therapeutic vaccines, and for diagnosis.

XX

XX Claim 1; Page 61-62; 118pp; English.

XX

CC The present sequence encodes a protein from Streptococcus pneumoniae. The

CC nucleic acid sequence encoding the Streptococcus pneumoniae protein can

CC be useful in vaccines for inducing protective antibodies against

CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.

CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid

CC are used to detect Streptococcus infection (by usual hybridisation or

CC amplification methods), also for isolating Streptococcus genes or their

CC allelic variants. The protein can be used similarly to detect specific

CC antibodies in standard immunoassays, especially for diagnosing or

CC monitoring infections. Antibodies which bind the protein are used to

CC detect corresponding antigens, to purify the protein and for passive

CC immunisation (optionally coupled to a toxin). Vaccines are administered,

CC e.g. by injection, orally or through the skin, typically at 0.01-1000

CC (especially 10-300) mu g/ml per dose

XX

SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;

Query Match 2.2%; Score 53; DB 2; Length 2290;

Best Local Similarity 100.0%; Pred. No. 6.8e-16;

Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATTTGGTACCGATTCAAG 1035

DB 944 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATTTGGTACCGATTCAAG 996

RESULT 15

ABQ84824

ID ABQ84824 standard; DNA; 2290 BP.

XX

AC ABQ84824;

XX

DT 04-SEP-2002 (first entry)

XX

DE S. pneumoniae SP042 nucleotide sequence SEQ ID NO:65.

XX

KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;

KW antibacterial; Streptococcal infection; detection; gene; ds.

XX

OS Streptococcus pneumoniae.

XX

PN US2002061545-A1.

XX

PD 23-MAY-2002.

XX

PF 22-JAN-2001; 2001US-00765272.

XX

PR 30-OCT-1997; 97US-00961083.

XX

PA (CHOI/) CHOI G H.

PA (KUNS/) KUNSCH C A.

PA (BARA/) BARASH S C.

PA (DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.

PA (FANN/) FANNON M R.

PA (ROSE/) ROSEN C A.

XX

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

PI Rosen CA;

XX

XX WPI; 2002-479261/51.

DR P-PSDB; ABP54589.

XX

PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus

PT and for preventing or attenuating disease caused by Streptococcus

PT infection.

XX

PS Claim 1; Page 28-29; 70pp; English.

XX WPI; 2002-122272/16.
DR P-PSDB; AAU75933.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
PT polypeptides, useful as vaccine components for treating or preventing
PT streptococcal infections such as otitis media, meningitis, and
PT bacteremia.
XX
XX Disclosure; Fig 4; 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90% identity
CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteremia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novicardiae or
CC Staphylococcus aureus) in an individual susceptible to the infection. A
CC polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence encodes the
CC Streptococcus pneumoniae protein BVH-11, used to create the antigenic
CC peptides described in the method of the invention
XX
SQ Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;

Query Match 6.2%; Score 148; DB 6; Length 2647;
Best Local Similarity 100.0%; Pred. No. 4.9e-64;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 699 CAAGACAACTGGGTACCTTCTGTAGCAATCCAGGAATCAAAATCACTAACCAAGCA 758
DB 802 CAAGACAACTGGGTACCTTCTGTAGCAATCCAGGAATCAAAATCACTAACCAAGCA 861

QY 759 ACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATCTCTTTGAAAC 818
DB 862 ACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATCTCTTTGAAAC 921

QY 819 AGCTCTACAACTGCCCTTTGAGTCAACG 846
DB 922 AGCTCTACAACTGCCCTTTGAGTCAACG 949

RESULT 12
AAA08556
ID AAA08556 standard; DNA; 504 BP.
XX
AC AAA08556;
XX
XX 19-JUL-2000 (first entry)
XX
DE S. pneumoniae 20 kDa human C3-degrading protein coding sequence.
XX
XX Human C3-degrading protein; 20 kDa; immunostimulatory; vaccine;
KW inhibitor; inflammation; organ rejection; xenotransplantation; ss.
XX
XX Streptococcus pneumoniae.
OS
XX WO200017370-A1.
FN
XX 30-MAR-2000.
PD
XX 24-SEP-1999; 99WO-US022362.
PF
XX 24-SEP-1998; 98US-0101736P.
PR

PR 31-MAR-1999; 99US-00283094.
XX (MINU) UNIV MINNESOTA.
PA (AMCY) AMERICAN CYANAMID CO.
XX
XX Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;
PI WPI; 2000-283594/24.
XX P-PSDB; AAY91938.
DR
XX Isolated polypeptide is used to stimulate immune system and immunize or
PT treat a mammalian subject against Streptococcus pneumoniae infection or
PT colonization.
XX
PS Claim 1; Page 53; 63pp; English.
XX
CC The present sequence, isolated from Streptococcus pneumoniae, encodes a
CC human C3-degrading protein (see AAY91938) of about 20 kDa. This sequence
CC may be part of a larger open reading frame (see AAA08557) which encodes
CC an approximately 92 kDa protein also having human C3-degrading activity.
CC The DNA sequences can be used for producing an immune response to
CC Streptococcus pneumoniae in a mammal. Antibodies against the proteins can
CC be used to inhibit S. pneumoniae-mediated C3 degradation. C3-mediated
CC inflammation and rejection in xenotransplantation can be inhibited by
CC expressing the nucleic acid sequences on the surface of an organ of an
CC animal. In particular, the polypeptides are useful for stimulating the
CC immune system and are effective to immunize or treat a mammalian subject
CC against Streptococcus pneumoniae infection or colonization
XX
SQ Sequence 504 BP; 153 A; 125 C; 98 G; 128 T; 0 U; 0 Other;

Query Match 2.2%; Score 53; DB 3; Length 504;
Best Local Similarity 100.0%; Pred. No. 6.4e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAAG 1035
DB 184 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAAG 236

RESULT 13
AAX25394
ID AAX25394 standard; DNA; 2163 BP.
XX
AC AAX25394;
XX
XX 19-JUL-1999 (first entry)
XX
XX Streptococcus pneumoniae complement C3-degrading protease DNA.
DE
XX Human complement C3-degrading protease; vaccine; infection; meningitis;
KW pneumonia; xenotransplantation; transplant rejection; inflammation; ds.
XX
XX Streptococcus pneumoniae.
OS
XX WO9915675-A1.
FN
XX 01-APR-1999.
PD
XX 24-SEP-1998; 98WO-US020186.
PF
XX 24-SEP-1997; 97US-0059907P.
PR (MINU) UNIV MINNESOTA.
PA (AMCY) AMERICAN CYANAMID CO.
XX
XX Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;
PI WPI; 1999-254719/21.
DR P-PSDB; AAY05753.
XX
XX New isolated human complement C3-degrading proteinase.
PT

```
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA001218.
XX PR 23-DEC-1998; 98US-0113800P.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX DR P-PSDB; AAB12716.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteremia and/or pneumonia.
XX PS Example 2; Fig 3; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
XX CC bactericidal activity. The nucleic acids, encoding the protein antigens,
XX CC may be used for the recombinant production of the proteins they encode.
XX CC The protein antigens may then be used as vaccines for the prevention and
XX CC treatment of Streptococcal infections in mammals (especially humans)
XX CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
XX CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
XX CC antigen
XX SQ Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 U; 0 Other;
XX
Query Match 6.2%; Score 148; DB 3; Length 2523;
Best Local Similarity 100.0%; Pred. No. 4.9e-64;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 699 CAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATCAAAATAGTATGATCTCTTGAAC 758
Db 758 CAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATCAAAATAGTATGATCTCTTGAAC 817
QY 759 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTGAAC 818
Db 818 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTGAAC 877
QY 819 AGCTCTACAAACTGCCTTTGAGTCAACG 846
Db 878 AGCTCTACAAACTGCCTTTGAGTCAACG 905
XX
RESULT 10
AAA65736
ID AAA65736 standard; DNA; 2647 BP.
XX AC AAA65736;
XX DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:12.
XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KW otitis media; pneumonia; immunisation; bactericidal; ds.
XX OS Streptococcus pneumoniae.
XX PN WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA001218.
XX PR 23-DEC-1998; 98US-0113800P.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX DR P-PSDB; AAB12716.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteremia and/or pneumonia.
XX PS Example 2; Fig 3; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
XX CC bactericidal activity. The nucleic acids, encoding the protein antigens,
XX CC may be used for the recombinant production of the proteins they encode.
XX CC The protein antigens may then be used as vaccines for the prevention and
XX CC treatment of Streptococcal infections in mammals (especially humans)
XX CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
XX CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
XX CC antigen
XX SQ Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 U; 0 Other;
XX
Query Match 6.2%; Score 148; DB 3; Length 2523;
Best Local Similarity 100.0%; Pred. No. 4.9e-64;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 699 CAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATCAAAATAGTATGATCTCTTGAAC 758
Db 758 CAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATCAAAATAGTATGATCTCTTGAAC 817
QY 759 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTGAAC 818
Db 818 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTGAAC 877
QY 819 AGCTCTACAAACTGCCTTTGAGTCAACG 846
Db 878 AGCTCTACAAACTGCCTTTGAGTCAACG 905
XX
RESULT 11
ABK15103
ID ABK15103 standard; DNA; 2647 BP.
XX AC ABK15103;
XX DT 08-MAY-2002 (first entry)
XX DE DNA encoding Streptococcus pneumoniae BVH-11.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
XX KW Streptococcal bacterial infection; gene; ds.
XX OS Streptococcus pneumoniae.
XX FH Key Location/Qualifiers
XX FT CDS 45..2567
XX FT /tag= a
XX FT /product= "BVH-11"
XX FT /note= "The gene is flanked by sequences from the vector
XX FT SP64, no information on which is given in the
XX FT specification"
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA000908.
XX PR 20-JUN-2000; 2000US-0212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
```

```
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX DR WPI; 2000-452397/39.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteremia and/or pneumonia.
XX PS Example 6; Fig 15; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
XX CC bactericidal activity. The nucleic acids, encoding the protein antigens,
XX CC may be used for the recombinant production of the proteins they encode.
XX CC The protein antigens may then be used as vaccines for the prevention and
XX CC treatment of Streptococcal infections in mammals (especially humans)
XX CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
XX CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
XX CC antigen
XX SQ Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;
XX
Query Match 6.2%; Score 148; DB 3; Length 2647;
Best Local Similarity 100.0%; Pred. No. 4.9e-64;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 699 CAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATCAAAATAGTATGATCTCTTGAAC 758
Db 802 CAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATCAAAATAGTATGATCTCTTGAAC 861
QY 759 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTGAAC 818
Db 862 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTGAAC 921
QY 819 AGCTCTACAAACTGCCTTTGAGTCAACG 846
Db 922 AGCTCTACAAACTGCCTTTGAGTCAACG 949
XX
RESULT 11
ABK15103
ID ABK15103 standard; DNA; 2647 BP.
XX AC ABK15103;
XX DT 08-MAY-2002 (first entry)
XX DE DNA encoding Streptococcus pneumoniae BVH-11.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
XX KW Streptococcal bacterial infection; gene; ds.
XX OS Streptococcus pneumoniae.
XX FH Key Location/Qualifiers
XX FT CDS 45..2567
XX FT /tag= a
XX FT /product= "BVH-11"
XX FT /note= "The gene is flanked by sequences from the vector
XX FT SP64, no information on which is given in the
XX FT specification"
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA000908.
XX PR 20-JUN-2000; 2000US-0212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
```

481 AGGAGCTATTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
Db
6470 AGGAGCTATTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 6411
Qy
541 TGGTGATGCTTATATCGTTCTCATGGAGATCATPACCATTAACATTCCTAAGAAATGAGTT 600
Db
6410 TGGTGATGCTTATATCGTTCTCATGGAGATCATPACCATTAACATTCCTAAGAAATGAGTT 6351
Qy
601 ATCAGCTAGCAGATGGCTGCTGCGAAGCCCTTCTATCTGGTCGAGGAATCTGTCAAA 660
Db
6350 ATCAGCTAGCAGATGGCTGCTGCGAAGCCCTTCTATCTGGTCGAGGAATCTGTCAAA 6291
Qy
661 TTCAAGAACCTTATCGCCGACAAAATAGCGATAACACTTTCAAGAACAAACTGGGTACCTTC 720
Db
6290 TTCAAGAACCTTATCGCCGACAAAATAGCGATAACACTTTCAAGAACAAACTGGGTACCTTC 6231
Qy
721 TGTAAAGCAATCCAGGAATCAAAATATCTAACAAGCAACAAACAGCACTAACAAGTCA 780
Db
6230 TGTAAAGCAATCCAGGAATCAAAATATCTAACAAGCAACAAACAGCACTAACAAGTCA 6171
Qy
781 AGCAAGTCAAAAGTAAATGACATTTAGTCTCTTGAACAGCTCTTACAACTGCCTTTGAG 840
Db
6170 AGCAAGTCAAAAGTAAATGACATTTAGTCTCTTGAACAGCTCTTACAACTGCCTTTGAG 6111
Qy
841 TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 900
Db
6110 TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC 6051
Qy
901 AGCTAGAGGTGTGAGTGCACACGAGATCATTAACACTTATCCCTTACTCTCAAAAT 960
Db
6050 AGCTAGAGGTGTGAGTGCACACGAGATCATTAACACTTATCCCTTACTCTCAAAAT 5991
Qy
961 GTCTCAATTTGAAGAACAAATGCTGCTATTATTTCCCTTGGTTATCGTTCAAACTTTG 1020
Db
5990 GTCTCAATTTGAAGAACAAATGCTGCTATTATTTCCCTTGGTTATCGTTCAAACTTTG 5931
Qy
1021 GGTACCAATTTCAAGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1080
Db
5930 GGTACCAATTTCAAGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 5871
Qy
1081 CCGGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGGT 1140
Db
5870 CCGGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGGT 5811
Qy
1141 ACGBAAAGTTCGGGAGGATATGATTCGAGAGAAAGGCACTCTCTGTTATGCTTTGTC 1200
Db
5810 ACGBAAAGTTCGGGAGGATATGATTCGAGAGAAAGGCACTCTCTGTTATGCTTTGTC 5751
Qy
1201 GAAAGATTTACCATCTGAAACTGTTTAAATAATCTTGAAGCAAGTTATCAAAACAAAGAG 1260
Db
5750 GAAAGATTTACCATCTGAAACTGTTTAAATAATCTTGAAGCAAGTTATCAAAACAAAGAG 5691
Qy
1261 TGTTCACACACTTTAATCTGTTAAATAATGTTGCTCTCTGACCAAGAAATTTTA 1320
Db
5690 TGTTCACACACTTTAATCTGTTAAATAATGTTGCTCTCTGACCAAGAAATTTTA 5631
Qy
1321 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAATAAGGGTCG 1380
Db
5630 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAATAAGGGTCG 5571
Qy
1381 TAAATCTGATTTCCAGCCTTAGACAAATTTATGAAAGCCTTTGAATGATGATCGACTAA 1440
Db
5570 TAAATCTGATTTCCAGCCTTAGACAAATTTATGAAAGCCTTTGAATGATGATCGACTAA 5511
Qy
1441 TAAAGAAAATTTGGTAGATGATTTATTTGGCATCTCTAGACCAATTAACCATCCAGAGG 1500
Db
5510 TAAAGAAAATTTGGTAGATGATTTATTTGGCATCTCTAGACCAATTAACCATCCAGAGG 5451
Qy
1501 ACTTCGGCAACCAAAATCTCAAAATTTAGTATGATGAAAGCAAGTTCGTTATGCTCAAT 1560
Db
5450 ACTTCGGCAACCAAAATCTCAAAATTTAGTATGATGAAAGCAAGTTCGTTATGCTCAAT 5391
Qy
1561 AGCTGATAAGTATACAAAGTCAGATGGTTATATTTTGTATGAACATGATATAATCAGTGA 1620

Db
5390 AGCTGATAAGTATACAAAGTCAGATGGTTACATTTTGTATGAACATGATATAATCAGTGA 5331
Qy
1621 TGAAGGAGATGATATGTAACGCCTCATATGGGCCATAGTCACTGGATTTGAAAAGATAG 1680
Db
5330 TGAAGGAGATGATATGTAACGCCTCATATGGGCCATAGTCACTGGATTTGAAAAGATAG 5271
Qy
1681 CCTTCTCATAGGAAAAGTTTGCAGCTCAAGCCTTATCTAAAGAAAAGGTTATCCTACC 1740
Db
5270 CCTTCTCATAGGAAAAGTTTGCAGCTCAAGCCTTATCTAAAGAAAAGGTTATCCTACC 5211
Qy
1741 TCCATCTCCAGACGAGATGTTTAAAGCAAATCCAACTGAGATAGTCAGACAGCTATTTA 1800
Db
5210 TCCATCTCCAGACGAGATGTTTAAAGCAAATCCAACTGAGATAGTCAGACAGCTATTTA 5151
Qy
1801 CAATCGTGTGAAGGGGAAAACGNAATCCACTGTTTCGACTTCATATATGGTTGAGCA 1860
Db
5150 CAATCGTGTGAAGGGGAAAACGNAATCCACTGTTTCGACTTCATATATGGTTGAGCA 5091
Qy
1861 TACAGTTTCAGGTTTAAAGACGGTAAATTTGATTTTCTCATTAAGGATCATTAACATAAT 1920
Db
5090 TACAGTTTCAGGTTTAAAGACGGTAAATTTGATTTTCTCATTAAGGATCATTAACATAAT 5031
Qy
1921 TAAATTTGCTTGGTTTGTATGATCAACATACAAAGCTCAAAATGGCTATACCTTGAAGA 1980
Db
5030 TAAATTTGCTTGGTTTGTATGATCAACATACAAAGCTCAAAATGGCTATACCTTGAAGA 4971
Qy
1981 TTTGTTTCGAGGATTAAGTACTAGTAGAACACCCCTGACGAACCTCCACATTTCTAATGA 2040
Db
4970 TTTGTTTCGAGGATTAAGTACTAGTAGAACACCCCTGACGAACCTCCACATTTCTAATGA 4911
Qy
2041 TGGATGGGGCAATGCGAGTGAAGTCTGTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
Db
4910 TGGATGGGGCAATGCGAGTGAAGTCTGTAGGCAAGAAAGACCAAGTGAAGATCCAAA 4851
Qy
2101 TAAGAACTTCAAAGCGGATGAAGCCAGTGAAGAAAACACTGCTGAGCCAGAAAGTCCC 2160
Db
4850 TAAGAACTTCAAAGCGGATGAAGCCAGTGAAGAAAACACTGCTGAGCCAGAAAGTCCC 4791
Qy
2161 TCAAGTAGAGACTGAAAAGTAGAGCCCACTCAAGAACGAGATTTGCTTGGCAA 2220
Db
4790 TCAAGTAGAGACTGAAAAGTAGAGCCCACTCAAGAACGAGATTTGCTTGGCAA 4731
Qy
2221 AGTAACGATTTCTAGTCTGAAAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
Db
4730 AGTAACGATTTCTAGTCTGAAAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 4671
Qy
2281 TTTGACTCTTCAAATTTATGATTAACAATAGTATATGCGAAGCAGAAAATTTACTTGC 2340
Db
4670 TTTGACTCTTCAAATTTATGATTAACAATAGTATATGCGAAGCAGAAAATTTACTTGC 4611
Qy
2341 GTTGTAAAGGAGTAAATCTTCTATCTGTAAGTAAAGGAAAATAAAC 2389
Db
4610 GTTGTAAAGGAGTAAATCTTCTATCTGTAAGTAAAGGAAAATAAAC 4562

RESULT 9

AAA65731

ID AAA65731 standard; DNA; 2523 BP.

XX

AC AAA65731;

XX

DT 21-NOV-2000 (first entry)

XX

DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.

XX

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia; otitis media; pneumonia; immunisation; bactericidal; ds.

XX

OS Streptococcus pneumoniae.

XX

PN WO200039299-A2.

| | | | |
|----|------|--|------|
| Db | 4913 | TACAGTTGAGGTTAAACCGTAAATTTGATTATTTCTCATAGGATCAATACCATAATAT | 4972 |
| Qy | 1921 | TAAATTTGCTTGGTTTGATGATCAACATCAAAAGCTCCAAATGGCTATACCTTGGAGA | 1980 |
| Db | 4973 | TAAATTTGCTTGGTTTGATGATCAACATCAAAAGCTCCAAATGGCTATACCTTGGAGA | 5032 |
| Qy | 1981 | TTTCTTTGCGACGAGTAAAGTAACTAGTAAACACCCCTGACCAACCTGACCAATCTTAATGA | 2040 |
| Db | 5033 | TTTCTTTGCGACGAGTAAAGTAACTAGTAAACACCCCTGACCAACCTGACCAATCTTAATGA | 5092 |
| Qy | 2041 | TGGATGGGGCAATGCCAGTGAAGCTGTGTTAGGCAAGAACACACAGTGAAGATCCAAA | 2100 |
| Db | 5093 | TGGATGGGGCAATGCCAGTGAAGCTGTGTTAGGCAAGAACACACAGTGAAGATCCAAA | 5152 |
| Qy | 2101 | TAAAGACTTCAAGCGGATGAAGAGCCAGTGAAGAACACCTGCTGAGCCAGAGTCCC | 2160 |
| Db | 5153 | TAAAGACTTCAAGCGGATGAAGAGCCAGTGAAGAACACCTGCTGAGCCAGAGTCCC | 5212 |
| Qy | 2161 | TCAAGTAGAGACTGAAAAGTAGAAGCCCAACTCAAGAGACAGAAAGTTTGGCTTGGCAA | 2220 |
| Db | 5213 | TCAAGTAGAGACTGAAAAGTAGAAGCCCAACTCAAGAGACAGAAAGTTTGGCTTGGCAA | 5272 |
| Qy | 2221 | AGTAACGGAATCTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA | 2280 |
| Db | 5273 | AGTAACGGAATCTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA | 5332 |
| Qy | 2281 | TTTGACTCTTCAAAATTTAGTAAATCAATAGTATCATGTCAGAGCAGAAATTAATTTGC | 2340 |
| Db | 5333 | TTTGACTCTTCAAAATTTAGTAAATCAATAGTATCATGTCAGAGCAGAAATTAATTTGC | 5392 |
| Qy | 2341 | GTTCGTTAAAGGAAGTAACTCTCATCTGTAAGTAAGGAAAAAATAAAC | 2389 |
| Db | 5393 | GTTCGTTAAAGGAAGTAACTCTCATCTGTAAGTAAGGAAAAAATAAAC | 5441 |

RESULT 7
 ABS56454_09/c
 Continuation (10 of 22) of ABS56454 from base 900001 (Streptococcus pneumoniae type 4 s)
 WP Sequence split into 22 fragments LOCUS ABS56454 Accession ABS56454

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | ABS56454_00 | 1 | 110000 |
| WP | ABS56454_01 | 100001 | 210000 |
| WP | ABS56454_02 | 200001 | 310000 |
| WP | ABS56454_03 | 300001 | 410000 |
| WP | ABS56454_04 | 400001 | 510000 |
| WP | ABS56454_05 | 500001 | 610000 |
| WP | ABS56454_06 | 600001 | 710000 |
| WP | ABS56454_07 | 700001 | 810000 |
| WP | ABS56454_08 | 800001 | 910000 |
| WP | ABS56454_09 | 900001 | 1010000 |
| WP | ABS56454_10 | 1000001 | 1110000 |
| WP | ABS56454_11 | 1100001 | 1210000 |
| WP | ABS56454_12 | 1200001 | 1310000 |
| WP | ABS56454_13 | 1300001 | 1410000 |
| WP | ABS56454_14 | 1400001 | 1510000 |
| WP | ABS56454_15 | 1500001 | 1610000 |
| WP | ABS56454_16 | 1600001 | 1710000 |
| WP | ABS56454_17 | 1700001 | 1810000 |
| WP | ABS56454_18 | 1800001 | 1910000 |
| WP | ABS56454_19 | 1900001 | 2010000 |
| WP | ABS56454_20 | 2000001 | 2110000 |
| WP | ABS56454_21 | 2100001 | 2162598 |

Query Match 97.9%; Score 2338; DB 7; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

| | | | |
|----|--------|--|--------|
| Qy | 1 | TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACCGTTAAGGAAAAATAATCGTGTTCCTA | 60 |
| Db | 106950 | TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACCGTTAAGGAAAAATAATCGTGTTCCTA | 106891 |
| Qy | 61 | TATAGATGAAAAACAGCGACGCAAAAAACCGAGAAATTTGACTCTCGTATGAGTTAGCAA | 120 |

| | | | |
|----|--------|--|--------|
| Db | 106890 | TATAGATGAAAAACAGCGACGCAAAAAACCGAGAAATTTGACTCTCGTATGAGTTAGCAA | 106831 |
| Qy | 121 | GCCTCAAGGATCAATCTGAGCAAAATCGTCAATCAAGATACACAGCAAGCTATGTCCAC | 180 |
| Db | 106830 | GCCTCAAGGATCAATCTGAGCAAAATCGTCAATCAAGATACACAGCAAGCTATGTCCAC | 106771 |
| Qy | 181 | TTACATGCGGCACCACTATCAATTTTCAATATGTAAGGTTCTCTTATGACGCTATCATCAG | 240 |
| Db | 106770 | TTACATGCGGCACCACTATCAATTTTCAATATGTAAGGTTCTCTTATGACGCTATCATCAG | 106711 |
| Qy | 241 | TGAAGAAATTAATCTATGAAAGATCCAAATTAAGTAAAGATGAGGATATTTTAATGA | 300 |
| Db | 106710 | TGAAGAAATTAATCTATGAAAGATCCAAATTAAGTAAAGATGAGGATATTTTAATGA | 106651 |
| Qy | 301 | GGTCAAGGTTGGATATGTTATCAAGGTAGATGGAATAATTAATTTTACCTTTAAGGATGC | 360 |
| Db | 106650 | GGTCAAGGTTGGATATGTTATCAAGGTAGATGGAATAATTAATTTTACCTTTAAGGATGC | 106591 |
| Qy | 361 | TGCCACGCGGATAAACGTCCTGACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG | 420 |
| Db | 106590 | TGCCACGCGGATAAACGTCCTGACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG | 106531 |
| Qy | 421 | TCAACATCGTGAAGTGGAACTCCAGAAACGATGCTGCTGCTTGGCCTTGGCAGCTTCGCA | 480 |
| Db | 106530 | TCAACATCGTGAAGTGGAACTCCAGAAACGATGCTGCTGCTTGGCCTTGGCAGCTTCGCA | 106471 |
| Qy | 481 | AGGACGCTATACATGATGATGTTATCTTTAATGCTTCTGATATCATAGAGATAC | 540 |
| Db | 106470 | AGGACGCTATACATGATGATGTTATCTTTAATGCTTCTGATATCATAGAGATAC | 106411 |
| Qy | 541 | TGGTGATGCTTATATCGTTCTCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT | 600 |
| Db | 106410 | TGGTGATGCTTATATCGTTCTCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT | 106351 |
| Qy | 601 | ATCAGTAGAGGTTGGCTGCTGCAAGAGCCTTCTCTATCTGGTCAGAGAAATCTCTCAA | 660 |
| Db | 106350 | ATCAGTAGAGGTTGGCTGCTGCAAGAGCCTTCTCTATCTGGTCAGAGAAATCTCTCAA | 106291 |
| Qy | 661 | TTCAGAACCTTATCGCCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC | 720 |
| Db | 106290 | TTCAGAACCTTATCGCCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC | 106231 |
| Qy | 721 | TGTAAGCAATCCAGGAATACAAAATTAACAAGCAACACAGCAACACCTAACAGTCA | 780 |
| Db | 106230 | TGTAAGCAATCCAGGAATACAAAATTAACAAGCAACACAGCAACACCTAACAGTCA | 106171 |
| Qy | 781 | AGCAAGTCAAAAGTAAATGATGATGATCTCTTGAAGAGCTCTTCAAACTGCCCTTGG | 840 |
| Db | 106170 | AGCAAGTCAAAAGTAAATGATGATGATCTCTTGAAGAGCTCTTCAAACTGCCCTTGG | 106111 |
| Qy | 841 | TCAACGATGTAGAAATCTGATGGCTTGTCTTTCATCGACACAAATCAACAGTCGAAC | 900 |
| Db | 106110 | TCAACGATGTAGAAATCTGATGGCTTGTCTTTCATCGACACAAATCAACAGTCGAAC | 106051 |
| Qy | 901 | AGCTAGAGGTTGTCAGTGCCACACGAGATCAATTAACCATTTATCCCTTACTCTCAAT | 960 |
| Db | 106050 | AGCTAGAGGTTGTCAGTGCCACACGAGATCAATTAACCATTTATCCCTTACTCTCAAT | 105991 |
| Qy | 961 | GTCTGAATGGAAGAACGAATCGCTGATTTATTCCTTCTGTTATCGTTCAAACTTGG | 1020 |
| Db | 105990 | GTCTGAATGGAAGAACGAATCGCTGATTTATTCCTTCTGTTATCGTTCAAACTTGG | 105931 |
| Qy | 1021 | GGTACCAAGTCAAGGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTTAGTCCAG | 1080 |
| Db | 105930 | GGTACCAAGTCAAGGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTTAGTCCAG | 105871 |
| Qy | 1081 | CCCGCAACCTGCACCAAAATCTTTAAATAGACTCAAAATTTCTTTGGTTAGTCAGCTGT | 1140 |
| Db | 105870 | CCCGCAACCTGCACCAAAATCTTTAAATAGACTCAAAATTTCTTTGGTTAGTCAGCTGT | 105811 |
| Qy | 1141 | ACGAAAAGTTGGGAGAGATGATTCGAGAAAAGGCACTCTCTGTTATGCTTTTGC | 1200 |
| Db | 105810 | ACGAAAAGTTGGGAGAGATGATTCGAGAAAAGGCACTCTCTGTTATGCTTTTGC | 105751 |

probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
 CC 391, identifying members of the library which contain sequences that
 CC hybridise to the target sequence and isolating the nucleic acid molecules
 CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an
 CC organism, amplifying nucleic acid molecules whose nucleotide sequence is
 CC homologous to amplification primers derived from the fragment of the S.
 CC pneumoniae genome to prime the amplification and isolating the amplified
 CC sequences. The computer readable medium can be used in a computer-based
 CC system for identifying fragments of the S. pneumoniae genome of
 CC commercial importance, or expression modulating fragments of the S.
 CC pneumoniae genome. Products from the present invention can be used in
 CC diagnosis kits and assays, and pharmaceutical compositions and vaccines
 CC for S. pneumoniae
 XX
 SQ Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 0 U; 3 Other;

Query Match 97.9%; Score 2338; DB 2; Length 8195;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

| | | | |
|----|------|---|------|
| QY | 1 | TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATTAATCGTGTTCCTA | 60 |
| DB | 3053 | TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATTAATCGTGTTCCTA | 3112 |
| QY | 61 | TATAGATGGAAACAAACGACGCGCAAAAACGAGAAATTTGACTCTCTGATGAGGTTAGCAA | 120 |
| DB | 3113 | TATAGATGGAAACAAACGACGCGCAAAAACGAGAAATTTGACTCTCTGATGAGGTTAGCAA | 3172 |
| QY | 121 | CGGTGAAGGAATCAATCTGAGCAAAATCGTCAATCAAGATACAGACCAAGGCTATGTCAC | 180 |
| DB | 3173 | CGGTGAAGGAATCAATCTGAGCAAAATCGTCAATCAAGATACAGACCAAGGCTATGTCAC | 3232 |
| QY | 181 | TTCAATCGGCGACCACTATCATTAATGATGGAAGGTTCCCTTATGACGCTATCATCAG | 240 |
| DB | 3233 | TTCAATCGGCGACCACTATCATTAATGATGGAAGGTTCCCTTATGACGCTATCATCAG | 3292 |
| QY | 241 | TGAAGAATTAATCATGAAGATCAAACTATAAGCTAAAGATGAGGATATGTTAATGA | 300 |
| DB | 3293 | TGAAGAATTAATCATGAAGATCAAACTATAAGCTAAAGATGAGGATATGTTAATGA | 3352 |
| QY | 301 | GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGC | 360 |
| DB | 3353 | GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGC | 3412 |
| QY | 361 | TGCCCAACGCGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG | 420 |
| DB | 3413 | TGCCCAACGCGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG | 3472 |
| QY | 421 | TCAACATCGTGAAGTGGAACTCCAGAAACGATGCTGCTTGGCTTGGCAGCTTCGCA | 480 |
| DB | 3473 | TCAACATCGTGAAGTGGAACTCCAGAAACGATGCTGCTTGGCTTGGCAGCTTCGCA | 3532 |
| QY | 481 | AGGACGCTATCTACAGATGAGTGTATATCTTTAATGCTTCTGATATCATAGAGATAC | 540 |
| DB | 3533 | AGGACGCTATCTACAGATGAGTGTATATCTTTAATGCTTCTGATATCATAGAGATAC | 3592 |
| QY | 541 | TGCTGATGCTTATATCGTCTCATGAGATCATTAACCATTAATCTTCCCTAAGAAATGAGTT | 600 |
| DB | 3593 | TGCTGATGCTTATATCGTCTCATGAGATCATTAACCATTAATCTTCCCTAAGAAATGAGTT | 3652 |
| QY | 601 | ATCAGCTAGCGAGTTGGCTGCTCGAAGCCTTCTTATCTGCTGCGAGAAATCTGTCAA | 660 |
| DB | 3653 | ATCAGCTAGCGAGTTGGCTGCTCGAAGCCTTCTTATCTGCTGCGAGAAATCTGTCAA | 3712 |
| QY | 661 | TTCAAGAACCTATCGCGCAAAATAGCGATACACTTCAAGAACAAACCTGGGTACCTTC | 720 |
| DB | 3713 | TTCAAGAACCTATCGCGCAAAATAGCGATACACTTCAAGAACAAACCTGGGTACCTTC | 3772 |
| QY | 721 | TGTAAGCAATCCAGGAATCAAAATACAAACAAAGCAACCAAGCAACACTAACAGTCA | 780 |
| DB | 3773 | TGTAAGCAATCCAGGAATCAAAATACAAACAAAGCAACCAAGCAACACTAACAGTCA | 3832 |
| QY | 781 | AGCAAGTCAAGATGACATGATGATCTCTTTGAAACACAGCTCTACAAACCTGCTTTGAG | 840 |

| | | | |
|----|------|---|------|
| DB | 3833 | AGCAAGTCAAGATTAATGACATTTGATGCTCTTGAACACAGCTCTACAAACCTGCTTTGAG | 3892 |
| QY | 841 | TCAACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGCAAAATCAAGTCCGAAC | 900 |
| DB | 3893 | TCAACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGCAAAATCAAGTCCGAAC | 3952 |
| QY | 901 | AGCTAGAGGTTGCGAGTGCACACGAGAGATCAATACCATCTCATCCCTTACTCTCAAT | 960 |
| DB | 3953 | AGCTAGAGGTTGCGAGTGCACACGAGAGATCAATACCATCTCATCCCTTACTCTCAAT | 4012 |
| QY | 961 | GTCTGAATTCGAAGAACGAATCGTGTATATTCCCTTCCCTTGTATCGTTCAACCAATG | 1020 |
| DB | 4013 | GTCTGAATTCGAAGAACGAATCGTGTATATTCCCTTCCCTTGTATCGTTCAACCAATG | 4072 |
| QY | 1021 | GGTACCAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG | 1080 |
| DB | 4073 | GGTACCAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG | 4132 |
| QY | 1081 | CCGCGAACCTGACCAAAATCTTAAATAGACTCAATTTCTTTTGGTTAGTCAGCTGGT | 1140 |
| DB | 4133 | CCGCGAACCTGACCAAAATCTTAAATAGACTCAATTTCTTTTGGTTAGTCAGCTGGT | 4192 |
| QY | 1141 | ACGAAAGTTGGGGAAGGATATGTTTGAAGAAAAAGGSCATCTCTCGTTATGCTTTTGC | 1200 |
| DB | 4193 | ACGAAAGTTGGGGAAGGATATGTTTGAAGAAAAAGGSCATCTCTCGTTATGCTTTTGC | 4252 |
| QY | 1201 | GAAAGATTTACCATCTGAAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAGAGAG | 1260 |
| DB | 4253 | GAAAGATTTACCATCTGAAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAGAGAG | 4312 |
| QY | 1261 | TGTTTCCACACACTTAACTGCTAAGAAAGAAATGTTGCTCTCGTGACCAAGAAATTTA | 1320 |
| DB | 4313 | TGTTTCCACACACTTAACTGCTAAGAAAGAAATGTTGCTCTCGTGACCAAGAAATTTA | 4372 |
| QY | 1321 | TGATAAAGCATATAATCTGTTAACTGAGGCTATAAAGCCCTCTTTGNAATTAAGGCTCG | 1380 |
| DB | 4373 | TGATAAAGCATATAATCTGTTAACTGAGGCTATAAAGCCCTCTTTGNAATTAAGGCTCG | 4432 |
| QY | 1381 | TAATTCGTATTTCCAGCCCTTAGACAAATATTTAGAACCGCTTGAATGATGAACTAA | 1440 |
| DB | 4433 | TAATTCGTATTTCCAGCCCTTAGACAAATATTTAGAACCGCTTGAATGATGAACTAA | 4492 |
| QY | 1441 | TAAAGAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTAACCATCCAGCG | 1500 |
| DB | 4493 | TAAAGAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTAACCATCCAGCG | 4552 |
| QY | 1501 | ACTTGGCAACCAAAATCTCAAATTCAGTATATCTGAAGACGAAATTCGTATTGCTCAAT | 1560 |
| DB | 4553 | ACTTGGCAACCAAAATCTCAAATTCAGTATATCTGAAGACGAAATTCGTATTGCTCAAT | 4612 |
| QY | 1561 | AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTGATGAACATGATATATAGTGA | 1620 |
| DB | 4613 | AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTGATGAACATGATATATAGTGA | 4672 |
| QY | 1621 | TGAAGAGATGATATGTAACCCCATATGCGCCATAGTCTCACTGATTTGGAAAGATAG | 1680 |
| DB | 4673 | TGAAGAGATGATATGTAACCCCATATGCGCCATAGTCTCACTGATTTGGAAAGATAG | 4732 |
| QY | 1681 | CTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCCTATATACTAAAGAAAAAGGTTCTCTAC | 1740 |
| DB | 4733 | CTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCCTATATACTAAAGAAAAAGGTTCTCTAC | 4792 |
| QY | 1741 | TCCATCTCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTGCAGAGCTATTTTA | 1800 |
| DB | 4793 | TCCATCTCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTGCAGAGCTATTTTA | 4852 |
| QY | 1801 | CAATCGTGTGAAGGGAAGAAACGAAATTCACCTCGTTCCACTTCCATATATGTTGAGCA | 1860 |
| DB | 4853 | CAATCGTGTGAAGGGAAGAAACGAAATTCACCTCGTTCCACTTCCATATATGTTGAGCA | 4912 |
| QY | 1861 | TACAGTTGAGGTTAAAAACCGGTAATTTGATTTATCTCTATAAGGATCATTAACATAAT | 1920 |

QY 901 AGCTAGAGGCTGTGCGAGTGCACACGAGATCATTACCACCTTCATCCCTTACTCTCAAT 960
 Db 918 AGCTAGAGGCTGTGCGAGTGCACACGAGATCATTACCACCTTCATCCCTTACTCTCAAT 977
 QY 961 GTCTGAATTTGAAGAACGAATCGCTCGTATTTATTTCCCTTCTGTTATCGTTCAAAACCATTTG 1020
 Db 978 GTCTGAATTTGAAGAACGAATCGCTCGTATTTATTTCCCTTCTGTTATCGTTCAAAACCATTTG 1037
 QY 1021 GGTACAGATTCAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1080
 Db 1038 GGTACAGATTCAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1097
 QY 1081 CCCGCAACTGCACCAATCTTAAATAGACTCAAAATCTCTTGGTGTAGTCACTGCT 1140
 Db 1098 CCCGCAACTGCACCAATCTTAAATAGACTCAAAATCTCTTGGTGTAGTCACTGCT 1157
 QY 1141 ACGAAAGTTGGGGAAGGATGTATTCGAAGAAAGGCGATCTCTCGTTATGTCCTTGC 1200
 Db 1158 ACGAAAGTTGGGGAAGGATGTATTCGAAGAAAGGCGATCTCTCGTTATGTCCTTGC 1217
 QY 1201 GAAAGATTTACATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260
 Db 1218 GAAAGATTTACATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1277
 QY 1261 TGTTTCACACCTTTAACTGCTTAAAGAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1320
 Db 1278 TGTTTCACACCTTTAACTGCTTAAAGAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1337
 QY 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTGC 1380
 Db 1338 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTGC 1397
 QY 1381 TAAATCTGATTTCCAGCCTTAGACAAATTTAGAACCTTGAATGATGAATCCACTAA 1440
 Db 1398 TAAATCTGATTTCCAGCCTTAGACAAATTTAGAACCTTGAATGATGAATCCACTAA 1457
 QY 1441 TAAAGAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATTAACCATCCAGAGCG 1500
 Db 1458 TAAAGAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATTAACCATCCAGAGCG 1517
 QY 1501 ACTTGGCAAAACCAATTTCAAATTTGATGATATCTGAAGACGAAGTTCGTTATGCTCAAT 1560
 Db 1518 ACTTGGCAAAACCAATTTCAAATTTGATGATATCTGAAGACGAAGTTCGTTATGCTCAAT 1577
 QY 1561 ACTGATAGATATACAGCTGAGATGTTACATTTTGTGATGACATGATATTAATCAGTCA 1620
 Db 1578 ACTGATAGATATACAGCTGAGATGTTACATTTTGTGATGACATGATATTAATCAGTCA 1637
 QY 1621 TGAAGGAGATGCATATGTAAACGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1680
 Db 1638 TGAAGGAGATGCATATGTAAACGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1697
 QY 1681 CCTTTCTGATTAAGGAAAAAGTTGACGCTCAAGCCTATATAAGAAAAAGGTTATCTTACC 1740
 Db 1698 CCTTTCTGATTAAGGAAAAAGTTGACGCTCAAGCCTATATAAGAAAAAGGTTATCTTACC 1757
 QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCAACTGGAGTAGTGCAGCGCTATTTA 1800
 Db 1758 TCCATCTCCAGACGAGATGTTAAAGCAATCAACTGGAGTAGTGCAGCGCTATTTA 1817
 QY 1801 CAATCGTGTGAAGGGGAAAAAGCAATTCACCTCGTTCCGATTCCTCATATATGTTGAGCA 1860
 Db 1818 CAATCGTGTGAAGGGGAAAAAGCAATTCACCTCGTTCCGATTCCTCATATATGTTGAGCA 1877
 QY 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTTATCTCTCATAGGATCATTACCATATAT 1920
 Db 1878 TACAGTTGAGGTTAAAAACGGTAATTTGATTTATCTCTCATAGGATCATTACCATATAT 1937
 QY 1921 TAAATTTGCTGTTGATGATCACAATCAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
 Db 1938 TAAATTTGCTGTTGATGATCACAATCAAGCTCCAAATGGCTATACCTTTGGAAGA 1997
 QY 1981 TTTGTTTGCAGCATTAAGTACTACGTAGAACCCCTGAGAACGCTCCACATTTCTAATGA 2040

Db 1998 TTTGTTTGCAGCATTAAGTACTACGTAGAACCCCTGACGAACGTCACATTTCTAATGA 2057
 QY 2041 TGGATGGGCAATGCCAGTGCATGTGTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
 Db 2058 TGGATGGGCAATGCCAGTGCATGTGTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2117
 QY 2101 TAAAGAACTTCAAGCGGATGAAGAGCCAGTGTAGAGAAACACCTGCTGAGCCAGAGTCCC 2160
 Db 2118 TAAAGAACTTCAAGCGGATGAAGAGCCAGTGTAGAGAAACACCTGCTGAGCCAGAGTCCC 2177
 QY 2161 TCAAGTAGAGACTGAAAAAGTGAAGAGCCCACTCAAGAGCAGAGTTTTCCTTGGCA 2220
 Db 2178 TCAAGTAGAGACTGAAAAAGTGAAGAGCCCACTCAAGAGCAGAGTTTTCCTTGGCA 2237
 QY 2221 AGTACCGGATCTAGTCTGAAAGCCCAATGCAACAGAACTCTAGTGTGTTTACGAATAA 2280
 Db 2238 AGTACCGGATCTAGTCTGAAAGCCCAATGCAACAGAACTCTAGTGTGTTTACGAATAA 2297
 QY 2281 TTTGACTCTTCAAAATTTATGATTAACAATAGTATCATGCGAAGCAGAAAAATTTACTGC 2340
 Db 2298 TTTGACTCTTCAAAATTTATGATTAACAATAGTATCATGCGAAGCAGAAAAATTTACTGC 2357
 QY 2341 GTTGTAAAAAGGAAGTAATCTTCTGTAAGTAAAGGAAAAATAAAC 2389
 Db 2358 GTTGTAAAAAGGAAGTAATCTTCTGTAAGTAAAGGAAAAATAAAC 2406

RESULT 6
 AAV52227
 ID AAV52227 standard; DNA; 8195 BP.
 XX AC AAV52227;
 XX DT 23-OCT-1998 (first entry)
 XX DE Streptococcus pneumoniae genome fragment SEQ ID NO:94.
 XX KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX OS Streptococcus pneumoniae.
 XX PN WO9818931-A2.
 XX PD 07-MAY-1998.
 XX PF 30-OCT-1997; 97WO-US019588.
 XX PR 31-OCT-1996; 96US-0029960P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
 PI Dougherty BA;
 XX WPI; 1998-272225/24.
 XX DR Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
 XX Claim 1; Page 727-732; 1409pp; English.
 CC The present invention describes a computer readable medium which has the
 CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
 CC on it, or a representative fragment or a sequence at least 95% identical
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
 CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
 CC pneumoniae. The present invention also describes an isolated nucleic acid
 CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
 CC by a process comprising: (a) screening a genomic DNA library using as a

QY 2281 TTGACTCTTCAAAATTATGGATAACAATAGTATCATCGCAGACGAGAAAATTACTTGC 2340
 Db |||||
 QY 2340 TTGACTCTTCAAAATTATGGATAACAATAGTATCATCGCAGACGAGAAAATTACTTGC 2399
 Db |||||
 QY 2341 GTTGTAAAAGGAGTAAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
 Db |||||
 QY 2400 GTTGTAAAAGGAGTAAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2448
 Db |||||

RESULT 5
 ABX06886
 ID ABX06886 standard; DNA; 2406 BP.
 XX
 AC ABX06886;
 XX
 DT 27-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain coding region #1174.
 XX
 KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae; type 4 strain.
 XX
 PN W0200277021-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-IB002163.
 XX
 PR 27-MAR-2001; 2001GB-00007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masignani V, Tettelin H, Fraser C;
 XX
 DR WPI; 2003-040579/03.
 DR P-PSDB; ABU01598.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX
 PS Claim 6; SEQ ID NO 2347; 56pp; English.
 XX

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the Streptococcus pneumoniae type 4 strain genomic sequence appearing as ABS5454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a Streptococcus nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a Streptococcus pneumoniae bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying

CC immunodominant proteins. The present sequence is one of the 2489
 CC identified coding region from the genomic sequence. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 2406 BP; 834 A; 461 C; 490 G; 621 T; 0 U; 0 Other;
 Query Match 97.9%; Score 2338; DB 7; Length 2406;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGAAAAAATATCGTTTCTTA 60
 Db 18 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGAAAAAATATCGTTTCTTA 77
 QY 61 TATAGATGGAACCAACAGCGACGCAAAACCGAGAAATTTGACTCCTGATGAGGTAGCAA 120
 Db 78 TATAGATGGAACCAACAGCGACGCAAAACCGAGAAATTTGACTCCTGATGAGGTAGCAA 137
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
 Db 138 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 197
 QY 181 TTCACATGGCGACCACTATCATTTATCAATGTAAGTTCCTTATGACGCTATCATCAG 240
 Db 198 TTCACATGGCGACCACTATCATTTATCAATGTAAGTTCCTTATGACGCTATCATCAG 257
 QY 241 TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATCTTAATGA 300
 Db 258 TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATCTTAATGA 317
 QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAATACTATGTTTACCTTAAGGATGC 360
 Db 318 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAATACTATGTTTACCTTAAGGATGC 377
 QY 361 TGCCGACGGGATAACGTCGCTACAAAGAGGAAATCAATCGACAAACCAAGAGCATAG 420
 Db 378 TGCCGACGGGATAACGTCGCTACAAAGAGGAAATCAATCGACAAACCAAGAGCATAG 437
 QY 421 TCAACATCTGGAAGTGGAACTCCAAAGAAACGATGCTGCTGTTGCCCTTGCGACGTTGCGCA 480
 Db 438 TCAACATCTGGAAGTGGAACTCCAAAGAAACGATGCTGCTGTTGCCCTTGCGACGTTGCGCA 497
 QY 481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
 Db 498 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 557
 QY 541 TGTGATGCTTATATCGTTCCCTCATCGAGATCATTAACCATTCATTCCTAAGGAATGAGTT 600
 Db 558 TGTGATGCTTATATCGTTCCCTCATCGAGATCATTAACCATTCATTCCTAAGGAATGAGTT 617
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGCGAAGCCCTTCCCTATCTGGTCGAGGAAATCTGTCAA 660
 Db 618 ATCAGCTAGCGAGTTGGCTGCTGCGAAGCCCTTCCCTATCTGGTCGAGGAAATCTGTCAA 677
 QY 661 TTCAAGAACCTATCGCCGCAAAATAGCGATACACTTCAAGAAACAACTGGGTACCTTC 720
 Db 678 TTCAAGAACCTATCGCCGCAAAATAGCGATACACTTCAAGAAACAACTGGGTACCTTC 737
 QY 721 TGTAAACCAATCCAGGAACCTACAAATACTAAACAGCAACCAACCACTAACAGTCA 780
 Db 738 TGTAAACCAATCCAGGAACCTACAAATACTAAACAGCAACCAACCACTAACAGTCA 797
 QY 781 AGCAAGTCAAAGTAATGATGATGATGCTTCTTGAACAGCTCTACAACTGCTTTGAG 840
 Db 798 AGCAAGTCAAAGTAATGATGATGATGCTTCTTGAACAGCTCTACAACTGCTTTGAG 857
 QY 841 TCAACGACATGTAGAAATCTGATGGCTTCTTGTGATCCAGCAAAATCACAAGTCAAC 900
 Db 858 TCAACGACATGTAGAAATCTGATGGCTTCTTGTGATCCAGCAAAATCACAAGTCAAC 917

Db 120 TATAGTGGAAAAAAGGCGACGCAAAAAACGAGAAATTTGACTCTCTGATGAGTTAGCAA 179
Qy 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATACAGACCAAGGCTATGTCAC 180
Db 180 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATACAGACCAAGGCTATGTCAC 239
Qy 181 TTCACATGGCGACCACTATCATATTATTAACAATGGTAAGGTTCTTATGACGCTATCATCAG 240
Db 240 TTCACATGGCGACCACTATCATATTATTAACAATGGTAAGGTTCTTATGACGCTATCATCAG 299
Qy 241 TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTGTAATGA 300
Db 300 TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTGTAATGA 359
Qy 301 GGTCAAGGGTGGATGTTATCAAGGTAGATGGAATAATACCTATGTTTACCTTAAAGGATGC 360
Db 360 GGTCAAGGGTGGATGTTATCAAGGTAGATGGAATAATACCTATGTTTACCTTAAAGGATGC 419
Qy 361 TGCCACACGGGATACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
Db 420 TGCCACACGGGATACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 479
Qy 421 TCAACATCGTGAAGGTGAATCCCAAGAAACGATGGTGTGTCCTTGGCGACGTTGCGA 480
Db 480 TCAACATCGTGAAGGTGAATCCCAAGAAACGATGGTGTGTCCTTGGCGACGTTGCGA 539
Qy 481 AGGACGCTATACAGATGATGTTATATCTTTAATGCTTCGTATCATAGAGGATAC 540
Db 540 AGGACGCTATACAGATGATGTTATATCTTTAATGCTTCGTATCATAGAGGATAC 599
Qy 541 TGGTGATGCTTATATCGTTCCTCATCGGATCATTAACCATTAACATTCCTAAGAATGAGTT 600
Db 600 TGGTGATGCTTATATCGTTCCTCATCGGATCATTAACCATTAACATTCCTAAGAATGAGTT 659
Qy 601 ATCAGCTAGCGATGGTGTGCGAGAGCCTTCCTATCTGTCGAGGAAATCTGTCAA 660
Db 660 ATCAGCTAGCGATGGTGTGCGAGAGCCTTCCTATCTGTCGAGGAAATCTGTCAA 719
Qy 661 TTCAGAACCTATCCGCGACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
Db 720 TTCAGAACCTATCCGCGACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 779
Qy 721 TGTAAAGCAATCCAGGAATACAAATACATAACAAGCAACAAGCAACACTAACAGTCA 780
Db 780 TGTAAAGCAATCCAGGAATACAAATACATAACAAGCAACAAGCAACACTAACAGTCA 839
Qy 781 AGCAAGTCAAGATATGACATTTAGTCTCTTTGAAAACAGCTCTACAACTGCCCTTGAG 840
Db 840 AGCAAGTCAAGATATGACATTTAGTCTCTTTGAAAACAGCTCTACAACTGCCCTTGAG 899
Qy 841 TCAACGACATGTAGATCTGTGGCTTGTCTTTGATCCAGACAAATCACAGTCCGAC 900
Db 900 TCAACGACATGTAGATCTGTGGCTTGTCTTTGATCCAGACAAATCACAGTCCGAC 959
Qy 901 AGCTAGAGGTGTGAGTGCACACGGAGATCATTAACCACTTCATCCCTTACTCTCAAT 960
Db 960 AGCTAGAGGTGTGAGTGCACACGGAGATCATTAACCACTTCATCCCTTACTCTCAAT 1019
Qy 961 GHTCGAATTTGGAAGAACGATCGCTGTAATTTATCCCTTGTGTTATCGTTCAAAACCAT 1020
Db 1020 GTCTGAATTTGGAAGAACGATCGCTGTAATTTATCCCTTGTGTTATCGTTCAAAACCAT 1079
Qy 1021 GTTACAGATTTCAAGCCAGAACCAAGTCCACAAACGCTCCGGAACCTAGTCCAGG 1080
Db 1080 GTTACAGATTTCAAGCCAGAACCAAGTCCACAAACGCTCCGGAACCTAGTCCAGG 1139
Qy 1081 CCCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGTT 1140
Db 1140 CCCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGTT 1199
Qy 1141 ACGAAAGTTGGGAGGATATGTTATCGAAGAAAGGCACTCTCGTTATGCTCTTTCG 1200
Db 1200 ACGAAAGTTGGGAGGATATGTTATCGAAGAAAGGCACTCTCGTTATGCTCTTTCG 1259

Qy 1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 1260
Db 1260 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 1319
Qy 1261 TGTTTCAACACTTTAACTGCTTAAAGAAAGAAATTTGCTCTCTGAGCAAGAAATTTTA 1320
Db 1320 TGTTTCAACACTTTAACTGCTTAAAGAAAGAAATTTGCTCTCTGAGCAAGAAATTTTA 1379
Qy 1321 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTGTTGNAATAAGGGTGG 1380
Db 1380 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTGTTGNAATAAGGGTGG 1439
Qy 1381 TAAATTCGATTTCCAAAGCCTTAGACAAAATTTATTAGAAGCCTTGAATGATGAATCGACTAA 1440
Db 1440 TAAATTCGATTTCCAAAGCCTTAGACAAAATTTATTAGAAGCCTTGAATGATGAATCGACTAA 1499
Qy 1441 TAAAGAAATTTGGTAGATGATTTTGGCATTTCTAGCACCAATTCCTCATCCAGAGG 1500
Db 1500 TAAAGAAATTTGGTAGATGATTTTGGCATTTCTAGCACCAATTCCTCATCCAGAGG 1559
Qy 1501 ACTTGGCAAAACCAATTTCTCAAATTTAGTATGTAATGAAGCAAGTTTCGTTATGCTCAAT 1560
Db 1560 ACTTGGCAAAACCAATTTCTCAAATTTAGTATGTAATGAAGCAAGTTTCGTTATGCTCAAT 1619
Qy 1561 AGCTGATTAAGTATACAAAGCTCAGATGGTTTACATTTTTTGTGAACATGATATATCACTGA 1620
Db 1620 AGCTGATTAAGTATACAAAGCTCAGATGGTTTACATTTTTTGTGAACATGATATATCACTGA 1679
Qy 1621 TGAAGAGATGATATGTAACCCCTCATATGGGCCATAGTCTACTGGATTGGAAAAGATAG 1680
Db 1680 TGAAGAGATGATATGTAACCCCTCATATGGGCCATAGTCTACTGGATTGGAAAAGATAG 1739
Qy 1681 CCTTCTGATAGGAAAAGTTGCGAGCTCAAGCCTATATCTAAGAAAAAGGTATCCTACC 1740
Db 1740 CCTTCTGATAGGAAAAGTTGCGAGCTCAAGCCTATATCTAAGAAAAAGGTATCCTACC 1799
Qy 1741 TCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGTATTTTA 1800
Db 1800 TCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGTATTTTA 1859
Qy 1801 CAATCGTGTGAAAAGGGGAAAAACGAATTCCTCTGTTGAGTTCATATATGTTGGTGA 1860
Db 1860 CAATCGTGTGAAAAGGGGAAAAACGAATTCCTCTGTTGAGTTCATATATGTTGGTGA 1919
Qy 1861 TACAGTGTGAGGTTTAAAGCGGTAATTTGATTTATCTCTATAAGGATCATTAACATAAT 1920
Db 1920 TACAGTGTGAGGTTTAAAGCGGTAATTTGATTTATCTCTATAAGGATCATTAACATAAT 1979
Qy 1921 TAAATTTGCTTGTGATGATCACAATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
Db 1980 TAAATTTGCTTGTGATGATCACAATACAAAGCTCCAAATGGCTATACCTTGGAGA 2039
Qy 1981 TTTGTTTGGAGCATTAAGTACTAGTAGAACACCTGTGAGAACGCTGCACATTTAATGA 2040
Db 2040 TTTGTTTGGAGCATTAAGTACTAGTAGAACACCTGTGAGAACGCTGCACATTTAATGA 2099
Qy 2041 TGGATGGGGCAATGCCAGTGCAGTGTGTTAGCGAAGAAAGACCAAGTGAAGATCCAAA 2100
Db 2100 TGGATGGGGCAATGCCAGTGCAGTGTGTTAGCGAAGAAAGACCAAGTGAAGATCCAAA 2159
Qy 2101 TAAAGAACTTTCAAAGCGGATGGAAGCCAGTAGAGGAAAACCTGCTGAGCCAGAGTCCC 2160
Db 2160 TAAAGAACTTTCAAAGCGGATGGAAGCCAGTAGAGGAAAACCTGCTGAGCCAGAGTCCC 2219
Qy 2161 TCAAGTAGAGATGAAAAGTAGAGGCCCACTCAAGAGCAGAGATTTGCTTCGGA 2220
Db 2220 TCAAGTAGAGATGAAAAGTAGAGGCCCACTCAAGAGCAGAGATTTGCTTCGGA 2279
Qy 2221 AGTAAAGGATTTAGTCTGAAAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAATAA 2280
Db 2280 AGTAAAGGATTTAGTCTGAAAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAATAA 2339

QY 1261 TGTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
 Db 1261 TGTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
 QY 1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTGNAATAAAGGGTCG 1380
 Db 1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTGNAATAAAGGGTCG 1380
 QY 1381 TAATTCGTGATTCGAAGCCTTAGACAATATTATAGAACGCTTGAATGATGAATCGACTAA 1440
 Db 1381 TAATTCGTGATTCGAAGCCTTAGACAATATTATAGAACGCTTGAATGATGAATCGACTAA 1440
 QY 1441 TAAAGAAAAATTTGTAGATGATTAATTTGCAATTCCTAGCACCAATACCATCCAGAGCG 1500
 Db 1441 TAAAGAAAAATTTGTAGATGATTAATTTGCAATTCCTAGCACCAATACCATCCAGAGCG 1500
 QY 1501 ACTTGGCAAAACCAAAATTCCTCAAAATGAGTATACCTGAAGACGAAAGTTCGTATTCCTCAAT 1560
 Db 1501 ACTTGGCAAAACCAAAATTCCTCAAAATGAGTATACCTGAAGACGAAAGTTCGTATTCCTCAAT 1560
 QY 1561 AGCTGATAGATATACACAGCTCAGATGCTTACATTTTGTAGCAATCATATATCACTGA 1620
 Db 1561 AGCTGATAGATATACACAGCTCAGATGCTTACATTTTGTAGCAATCATATATCACTGA 1620
 QY 1621 TGAAGGAGATGCTATGTAACGCTCATATGGGCCATAGTCACTGGATTCGAAAAGATAG 1680
 Db 1621 TGAAGGAGATGCTATGTAACGCTCATATGGGCCATAGTCACTGGATTCGAAAAGATAG 1680
 QY 1681 CCTTCTGTAGTAAGAAAAAGTTGAGCTCAGCTCAAGCTTATCTAAAGAAAAAGTATCCTACC 1740
 Db 1681 CCTTCTGTAGTAAGAAAAAGTTGAGCTCAGCTCAAGCTTATCTAAAGAAAAAGTATCCTACC 1740
 QY 1741 TCCATCTCCAGACGCGATGTTAAAGCAATCCAACTGGAGATAGTCGACGACTATTTA 1800
 Db 1741 TCCATCTCCAGACGCGATGTTAAAGCAATCCAACTGGAGATAGTCGACGACTATTTA 1800
 QY 1801 CAATCGTGTGAAAGGGAAAAAGCAATTCACATCGTTCCGACTCCATATATGTTGAGCA 1860
 Db 1801 CAATCGTGTGAAAGGGAAAAAGCAATTCACATCGTTCCGACTCCATATATGTTGAGCA 1860
 QY 1861 TACAGTTGAGTTAAACCGGTATTTGATATTTCCTCATAGATCATTACCATATAT 1920
 Db 1861 TACAGTTGAGTTAAACCGGTATTTGATATTTCCTCATAGATCATTACCATATAT 1920
 QY 1921 TAAATTTGCTTTGTTGATCATCACATCAAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
 Db 1921 TAAATTTGCTTTGTTGATCATCACATCAAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
 QY 1981 TTTGTTTGGACGATTAAAGTACTACGTAGAACACCCCTGACGAACGTCCTCATTTCTAATGA 2040
 Db 1981 TTTGTTTGGACGATTAAAGTACTACGTAGAACACCCCTGACGAACGTCCTCATTTCTAATGA 2040
 QY 2041 TGATGGGGCAATGCCAGTGAGATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100
 Db 2041 TGATGGGGCAATGCCAGTGAGATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100
 QY 2101 TAAGAACTTCAAAGCGATCAAGACGAGTACGAGGAAACACCTGCTGAGCCAGATCCC 2160
 Db 2101 TAAGAACTTCAAAGCGATCAAGACGAGTACGAGGAAACACCTGCTGAGCCAGATCCC 2160
 QY 2161 TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAAGAGCAGAAAGTTTGTGCTGCGAA 2220
 Db 2161 TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAAGAGCAGAAAGTTTGTGCTGCGAA 2220
 QY 2221 AGTAACGGATTCAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
 Db 2221 AGTAACGGATTCAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
 QY 2281 TTTGACTCTTCAAATATGATACCAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC 2340
 Db 2281 TTTGACTCTTCAAATATGATACCAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC 2340
 QY 2341 GTTGTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389

Db 2341 GTTGTAAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC 2389
 RESULT 4
 ID AAA47604 standard; DNA; 2451 BP.
 AC AAA47604;
 XX 20-OCT-2000 (first entry)
 XX Recombinant variant of Sp36 gene (Sp36A) of S. pneumoniae.
 XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; ds.
 XX Streptococcus pneumoniae.
 OS
 XX Key Location/Qualifiers
 FH 1. .2451
 CDS /*tag= a
 FT /product= "Sp36A polypeptide"
 XX
 PN WO200037105-A2.
 XX
 PD 29-JUN-2000.
 XX
 PF 21-DEC-1999; 99WO-US030390.
 XX
 PR 21-DEC-1998; 98US-0113048P.
 XX
 PA (MEDI-) MEDIMUNE INC.
 XX
 PI Johnson LS, Koenig S, Adamou JE;
 XX
 WPI; 2000-452129/39.
 DR P-PSDB; AAB01468.
 XX
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections, comprises
 PT Streptococcus pneumoniae proteins.
 XX
 PS Disclosure; Page 64-65; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved in
 CC the pathogenicity of Streptococcus pneumoniae, there still remains a need
 CC to identify polypeptides having epitopes in common from various strains
 CC of S. pneumoniae in order to utilize such polypeptides in vaccines to
 CC protect against a wide variety of S. pneumoniae. New vaccine compositions
 CC are described which comprise a Streptococcus pneumoniae polypeptide (or
 CC fragments) of 80 - 680 amino acids in length that comprise at least one
 CC histidine triad residue (HxxHxxH) or a coiled-coil region, or an antibody
 CC directed against these features. The vaccine is useful in protecting
 CC against infection by Streptococcus pneumoniae. The vaccine composition
 CC comprising antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal and
 CC bronchial infections
 XX
 SQ Sequence 2451 BP; 849 A; 467 C; 499 G; 635 T; 0 U; 1 Other;
 Query Match 100.0%; Score 2389; DB 3; Length 2451;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTCTTACGAGTTGGGACTGTATCAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTTA 60
 Db 60 TTCTTACGAGTTGGGACTGTATCAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTTA 119
 QY 61 TATAGATGAAAAACAGCGACGCAAAAAACGGAGAAATTTGACTCCTCATGAGTTAGCAA 120

QY 2341 GTTGTAAAGGAGTAATCCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389
 Db 2341 GTTGTAAAGGAGTAATCCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389

RESULT 3

AD45136
 ID AD45136 standard; DNA; 2389 BP.

XX AD45136;

DT 18-DEC-2003 (first entry)

XX S. pneumoniae DNA encoding antigen SP036.

XX Antigen; ds; bacterial infection; vaccine; pneumonia; antibacterial.

XX Streptococcus pneumoniae.

XX US6573082-B1.

XX 03-JUN-2003.

XX 28-MAR-2000; 2000US-00536784.

XX 31-OCT-1996; 96US-0029960P.

XX 30-OCT-1997; 97US-00961083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

XX Rosen CA;

XX WPI; 2003-764574/72.

XX P-PSDB; AD45137.

XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
 useful for producing vaccines for prevention or attenuation of infection
 by Streptococcus pneumoniae.

XX Example 1; SEQ ID NO 55; 58pp; English.

XX The invention relates to an isolated polynucleotide consisting of a
 Streptococcus pneumoniae nucleic acid (appearing as ABC45122 and encoding
 SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
 antigens. Also included are making a recombinant vector by inserting the
 nucleic acid into a vector, an isolated polynucleotide consisting of at
 least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
 recombinant host cell comprising the SP028 polynucleotide. The nucleic
 acids are useful as DNA vaccine against Streptococcus pneumoniae
 infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
 antigen nucleic acids are useful as probes for use in diagnostic methods
 for detecting S. pneumoniae gene expression. The present sequence encodes
 an S. pneumoniae antigenic protein.

XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;

Query Match 100.0%; Score 2389; DB 9; Length 2389;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAGGAAAAATAATCGGTTTCCTA 60
 Db 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAGGAAAAATAATCGGTTTCCTA 60

QY 61 TATAGATGAAACAAAGCGACGCCAAAAACCGAGAAATTGATCTCTGATGAGGTTAGCAA 120
 Db 61 TATAGATGAAACAAAGCGACGCCAAAAACCGAGAAATTGATCTCTGATGAGGTTAGCAA 120

QY 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAGATAACAGACCAAGGTTATGTCAC 180
 Db 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAGATAACAGACCAAGGTTATGTCAC 180

QY 181 TTCATATGCGACCACTATCATTTATTAACAATGGTAAGGTTCTTATGACGCTATCATCAG 240
 Db 181 TTCATATGCGACCACTATCATTTATTAACAATGGTAAGGTTCTTATGACGCTATCATCAG 240
 QY 241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAGATAGGATATTTGTTAATGA 300
 Db 241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAGATAGGATATTTGTTAATGA 300
 QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAATCAATCGACAAAAACAAGAGCATAG 420
 Db 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAATCAATCGACAAAAACAAGAGCATAG 420
 QY 421 TCAACATCGTGAAGGTGAACTCCCAAGAAACGATGGTCTGTGGCTTTGGCAAGTTCGCA 480
 Db 421 TCAACATCGTGAAGGTGAACTCCCAAGAAACGATGGTCTGTGGCTTTGGCAAGTTCGCA 480
 QY 481 AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 Db 481 AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 QY 541 TGGTGATGCTTATATCGTTTCTCATGAGATCATTAACCATTAACATTCCTAAGAATGAGTT 600
 Db 541 TGGTGATGCTTATATCGTTTCTCATGAGATCATTAACCATTAACATTCCTAAGAATGAGTT 600
 QY 601 ATCAGTAGGAGTTGGCTGTGCGAAGCCTTCTATCTGGTCAGGAAATCTGTCAAA 660
 Db 601 ATCAGTAGGAGTTGGCTGTGCGAAGCCTTCTATCTGGTCAGGAAATCTGTCAAA 660
 QY 661 TTCAGAACCTATCGCCGACAAATAGCGATAACAACTTCAAGAAACAACTGGTGCTTCTC 720
 Db 661 TTCAGAACCTATCGCCGACAAATAGCGATAACAACTTCAAGAAACAACTGGTGCTTCTC 720
 QY 721 TGTAAAGCAATCCAGAACTTACAAATATAACAACAAGCAACAAACAGCAACACTTAACAGTCA 780
 Db 721 TGTAAAGCAATCCAGAACTTACAAATATAACAACAAGCAACAAACAGCAACACTTAACAGTCA 780
 QY 781 AGCAAGTCAAGTAATGATGATGATCTCTTGAACAGCTCTTCAAACTGCCCTTTGAG 840
 Db 781 AGCAAGTCAAGTAATGATGATGATCTCTTGAACAGCTCTTCAAACTGCCCTTTGAG 840
 QY 841 TCAACGACATGATGAGTCTGTGCTTGTTCATCCAGCAACAAATCACAAGTTCGAAC 900
 Db 841 TCAACGACATGATGAGTCTGTGCTTGTTCATCCAGCAACAAATCACAAGTTCGAAC 900
 QY 901 AGCTAGAGTGTGTCAGTGCCACACGAGATCATTAACCACTTCACTCCCTTACTCTCAAT 960
 Db 901 AGCTAGAGTGTGTCAGTGCCACACGAGATCATTAACCACTTCACTCCCTTACTCTCAAT 960
 QY 961 GTCTGAATTGGAAGAACGAATCGCTGATTTATTCCTGTTATGTTTCTGTTTCAACCATG 1020
 Db 961 GTCTGAATTGGAAGAACGAATCGCTGATTTATTCCTGTTATGTTTCTGTTTCAACCATG 1020
 QY 1021 GGTACAGATTCAAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAG 1080
 Db 1021 GGTACAGATTCAAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAG 1080
 QY 1081 CCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGTGT 1140
 Db 1081 CCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGTGT 1140
 QY 1141 AGCAAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGCACTCTCGTTATGTTCTTGC 1200
 Db 1141 AGCAAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGCACTCTCGTTATGTTCTTGC 1200
 QY 1201 GAAAGATTACCATCTGAAACTGTTAAATACTCTGAAGCAAGTATCAAAAACAAGAGAG 1260
 Db 1201 GAAAGATTACCATCTGAAACTGTTAAATACTCTGAAGCAAGTATCAAAAACAAGAGAG 1260

Db 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 180
QY 181 TTGACATGGCAGCACATCATATTAATACAAATGTTAGGTTCTCTATGACGCTATCATCAG 240
Db 181 TTCATGCGGCACCACTATCATTAATTAACAATGGTAAGGTTCTCTATGACGCTATCATCAG 240
QY 241 TGAAGAATTACTCATGAAAGATCCAACTATATAAGCTAAAGATGAGGATATTTGTAATGA 300
Db 241 TGAAGAAATTACTCATGAAGATCCAACTATTAAGCTAAAGATGAGGATATTTGTAATGA 300
QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTTAAGGATGC 360
Db 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTTAAGGATGC 360
QY 361 TGCCCGCGGATACAGTCCGCTACAAAGAGCAATCAATCGACAAAAAACAAAGAGCATAG 420
Db 361 TGCCCGCGGATACAGTCCGCTACAAAGAGCAATCAATCGACAAAAAACAAAGAGCATAG 420
QY 421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTCTGCTTGGCAGCTTCGCA 480
Db 421 TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTCTGCTTGGCAGCTTCGCA 480
QY 481 AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTCATATCATAGAGGATAC 540
Db 481 AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTCATATCATAGAGGATAC 540
QY 541 TGGTGATGCTTATATCGTTCTCATGAGATCATACCATACATTTCTTAAGATGAGTT 600
Db 541 TGGTGATGCTTATATCGTTCTCATGAGATCATACCATACATTTCTTAAGATGAGTT 600
QY 601 ATCAGCTAGCGAGTGGCTGTCGAGAGCCCTCTATCTGGTCGAGAAATCTGTCAA 660
Db 601 ATCAGCTAGCGAGTGGCTGTCGAGAGCCCTCTATCTGGTCGAGAAATCTGTCAA 660
QY 661 TTCAAGAACCTATCGCGCAGAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
Db 661 TTCAAGAACCTATCGCGCAGAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
QY 721 TGTAAAGCAATCCAGGAATCAATATCTAACACAGCAACACACAGCAACTAACAGTCA 780
Db 721 TGTAAAGCAATCCAGGAATCAATATCTAACACAGCAACACACAGCAACTAACAGTCA 780
QY 781 AGCAAGTCAAAGTAAATGACATTTGATAGTCTCTTGAAACAGCTCTACAAACTGGCTTTGAG 840
Db 781 AGCAAGTCAAAGTAAATGACATTTGATAGTCTCTTGAAACAGCTCTACAAACTGGCTTTGAG 840
QY 841 TCAACGACATGTAGAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGTCAAC 900
Db 841 TCAACGACATGTAGAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGTCAAC 900
QY 901 AGCTAGAGGTGTTGCAAGTGCACACGAGATCATACCACTTCTATCCCTTACTCTCAAT 960
Db 901 AGCTAGAGGTGTTGCAAGTGCACACGAGATCATACCACTTCTATCCCTTACTCTCAAT 960
QY 961 GTCTGAATTTGGAAGAACGAATCGCTGTAATATTTCCCTTTGTTATCGTTCAAACTATG 1020
Db 961 GTCTGAATTTGGAAGAACGAATCGCTGTAATATTTCCCTTTGTTATCGTTCAAACTATG 1020
QY 1021 GGTACAGATTTCAAGGCCAGAAACCAAGTCCAAACGACCTCGGAACTTATGTCAGG 1080
Db 1021 GGTACAGATTTCAAGGCCAGAAACCAAGTCCAAACGACCTCGGAACTTATGTCAGG 1080
QY 1081 CCGCAACCTGCAACCAATCTTAAATPAGACTCAAATTTCTTTGTTAGTCAGCTGGT 1140
Db 1081 CCGCAACCTGCAACCAATCTTAAATPAGACTCAAATTTCTTTGTTAGTCAGCTGGT 1140
QY 1141 ACAGAAAGTTGGGAAGGATATGTTATTCGAAGAAAGGGGATCTCTCGTTATGCTTTTC 1200
Db 1141 ACAGAAAGTTGGGAAGGATATGTTATTCGAAGAAAGGGGATCTCTCGTTATGCTTTTC 1200
QY 1201 GAAAGATTTACCATCTGAAATCTGTTAAATCTTTGAAAGCAAGTTATCAAAAACAGAG 1260

Db 1201 GAAAGATTTACCATCTGAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 1260
QY 1261 TGTTTTCACACATTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTCACCAAGAAATTTTA 1320
Db 1261 TGTTTTCACACATTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTCACCAAGAAATTTTA 1320
QY 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTAATAAGGGTCG 1380
Db 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTAATAAGGGTCG 1380
QY 1381 TAAATTCGATTTCCAGCCCTTAGACAAATATTAGAACGCTTGAATGATGATCGACTAA 1440
Db 1381 TAAATTCGATTTCCAGCCCTTAGACAAATATTAGAACGCTTGAATGATGATCGACTAA 1440
QY 1441 TAAAGAAAAATTTGTTAGATGATTTATTGGCATTCTTAGCACCATTACCCATCCAGAGCG 1500
Db 1441 TAAAGAAAAATTTGTTAGATGATTTATTGGCATTCTTAGCACCATTACCCATCCAGAGCG 1500
QY 1501 ACTTGGCAAAACCAAAATTTCTCAAAATTTAGTATCTGAAGAGCAAGTTCTGTTCTCAAT 1560
Db 1501 ACTTGGCAAAACCAAAATTTCTCAAAATTTAGTATCTGAAGAGCAAGTTCTGTTCTCAAT 1560
QY 1561 AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTGATGAACATGATATAACAGTGA 1620
Db 1561 AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTGATGAACATGATATAACAGTGA 1620
QY 1621 TGAAGGATGCTATATGTTAAAGGCTCATATGGGCTATGCTGGAATGGAAAAAGATAG 1680
Db 1621 TGAAGGATGCTATATGTTAAAGGCTCATATGGGCTATGCTGGAATGGAAAAAGATAG 1680
QY 1681 CCTTCTGATAAGCAAAAGTTGTCAGCTCAAGGCTATCTAAAGAAAAAGGATCTCTTACC 1740
Db 1681 CCTTCTGATAAGCAAAAGTTGTCAGCTCAAGGCTATCTAAAGAAAAAGGATCTCTTACC 1740
QY 1741 TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
Db 1741 TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
QY 1801 CAATCTGTTGAAAGGGGAAAAAGCAATTCACCTGTTGCACTTCCATATATGTTGAGCA 1860
Db 1801 CAATCTGTTGAAAGGGGAAAAAGCAATTCACCTGTTGCACTTCCATATATGTTGAGCA 1860
QY 1861 TACAGTTGAGGTTTAAAAACGTTAATTTCTCTATAAGGATCATTTACCATATAT 1920
Db 1861 TACAGTTGAGGTTTAAAAACGTTAATTTCTCTATAAGGATCATTTACCATATAT 1920
QY 1921 TAAATTTGCTTGGTTTGTATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
Db 1921 TAAATTTGCTTGGTTTGTATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
QY 1981 TTTGTTTGGACGATTAAGTACTACCTAGAACACCTTGACGAACGCTCCATTTCTAATGA 2040
Db 1981 TTTGTTTGGACGATTAAGTACTACCTAGAACACCTTGACGAACGCTCCATTTCTAATGA 2040
QY 2041 TGATGGGCAATTCGCAAGTGAAGCATGTTTAGCAAGAAAGACCAAGTGAAGATCCAAA 2100
Db 2041 TGATGGGCAATTCGCAAGTGAAGCATGTTTAGCAAGAAAGACCAAGTGAAGATCCAAA 2100
QY 2101 TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACCTGTGTAGCCAGCAAGTCCC 2160
Db 2101 TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACCTGTGTAGCCAGCAAGTCCC 2160
QY 2161 TCAAGTAGAGATGTAAGAGTGAAGCCCACTCAAGAGCAGAGATTTGTTGTCGAA 2220
Db 2161 TCAAGTAGAGATGTAAGAGTGAAGCCCACTCAAGAGCAGAGATTTGTTGTCGAA 2220
QY 2221 AGTAAACGATTTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
Db 2221 AGTAAACGATTTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
QY 2281 TTTGACTCTTCAAAATTTAGGATAACAATAGTATCATGGCAGAGCAGAAAAATTTACTTC 2340
Db 2281 TTTGACTCTTCAAAATTTAGGATAACAATAGTATCATGGCAGAGCAGAAAAATTTACTTC 2340


```
1261 TGTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
1261 TGTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
1321 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAAGGGTCG 1380
1321 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAAGGGTCG 1380
1381 TAAATTCGATTTTCCAAGCCTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 1440
1381 TAAATTCGATTTTCCAAGCCTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 1440
1441 TAAAGAAAATTTGGTAGATGATTTTATGCGCATTCCTAGCACCATTACCCATCCGAGGG 1500
1441 TAAAGAAAATTTGGTAGATGATTTTATGCGCATTCCTAGCACCATTACCCATCCGAGGG 1500
1501 ACTTCGCAAAACCAATTTCTCAATTTGAGTATATCTGAAGACGAATTTGTTATGCTCAAT 1560
1501 ACTTCGCAAAACCAATTTCTCAATTTGAGTATATCTGAAGACGAATTTGTTATGCTCAAT 1560
1561 AGCTGATAGTATAACAAGTCAGATGGTTACATTTTGTGTAACATGATATATCATGTA 1620
1561 AGCTGATAGTATAACAAGTCAGATGGTTACATTTTGTGTAACATGATATATCATGTA 1620
1621 TGAAGGAGTATATGTAACGCTCATATGGCCATAGTCTGATGGAAGATAG 1680
1621 TGAAGGAGTATATGTAACGCTCATATGGCCATAGTCTGATGGAAGATAG 1680
1681 CTTTCTGATAGGAAAAGTTGCGCTCAAGCTATATCTAAAGAAAAGGTATCTTACC 1740
1681 CTTTCTGATAGGAAAAGTTGCGCTCAAGCTATATCTAAAGAAAAGGTATCTTACC 1740
1741 TCCATCTCCAGACGAGATGTTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
1741 TCCATCTCCAGACGAGATGTTTAAAGCAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
1801 CAATCGTGTGAAGGGGAAAAGCAATTCACCTGTCGACTTCCATATATGTTGAGCA 1860
1801 CAATCGTGTGAAGGGGAAAAGCAATTCACCTGTCGACTTCCATATATGTTGAGCA 1860
1861 TACAGTTGAGTTTAAAGAAAGTAAATTTGATTTTCTCATAGGATCAATTTACCATATAT 1920
1861 TACAGTTGAGTTTAAAGAAAGTAAATTTGATTTTCTCATAGGATCAATTTACCATATAT 1920
1921 TAAATTTGCTTGGTTTGAATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
1921 TAAATTTGCTTGGTTTGAATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
1981 TTTGTTTGGAGATTAAGTACTACCTAGACACACCTGAGGACGCTCCATTTCTAATGA 2040
1981 TTTGTTTGGAGATTAAGTACTACCTAGACACACCTGAGGACGCTCCATTTCTAATGA 2040
2041 TGGATGGGCAATGCGAGTGGATGTTTGGTAAAGAAAGACACAGTGAAGATCCAA 2100
2041 TGGATGGGCAATGCGAGTGGATGTTTGGTAAAGAAAGACACAGTGAAGATCCAA 2100
2101 TAAGAACTTCAAAGCGGATGAAGCGGATGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
2101 TAAGAACTTCAAAGCGGATGAAGCGGATGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
2161 TCAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAGAGAGCAGAGTTTGTGTCGAA 2220
2161 TCAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAGAGAGCAGAGTTTGTGTCGAA 2220
2221 AGTAAAGGATTTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
2221 AGTAAAGGATTTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
2281 TTTGACTCTTCAATTTAGATACAAATAGTATCATGCGAGAGCAGAAAATTTACTTGC 2340
2281 TTTGACTCTTCAATTTAGATACAAATAGTATCATGCGAGAGCAGAAAATTTACTTGC 2340
```

2341 GTTGTAAAGGAGAGTAATCTCTTCATCTGTAGTAAGGAAAAATAAAC 2389
|||||
2341 GTTGTAAAGGAGAGTAATCTCTTCATCTGTAGTAAGGAAAAATAAAC 2389
|||||

RESULT 2
ABQ84819
ID ABQ84819 standard; DNA; 2389 BP.
XX ABQ84819;
XX
XX 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP036 nucleotide sequence SEQ ID NO:55.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection; gene; ds.
XX
OS Streptococcus pneumoniae.
XX
XX US2002061545-A1.
XX
XX 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-00765272.
XX
XX 30-OCT-1997; 97US-00961083.
XX
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.
XX (DOUG/) DOUGHERTY B.
XX (FANN/) FANNON M R.
XX (ROSE/) ROSEN C A.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX
XX WPI; 2002-479261/51.
XX P-PSDB; ABP54584.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX and for preventing or attenuating disease caused by Streptococcus
XX infection.
XX
XX Claim 1; Page 27; 70pp; English.
XX
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX pneumoniae antigens have antibacterial activity and can be used in
XX vaccines. The S. pneumoniae antigens can also be used to prevent or
XX attenuate a Streptococcal infection in an animal. The polynucleotides
XX encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX of S. pneumoniae ORFs (open reading frames) which are used in an example
XX from the present invention
XX
XX SQ Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;

Query Match 100.0%; Score 2389; DB 6; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTTAAAGAAAATAATCGTGTTCCTTA 60
DB 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTTAAAGAAAATAATCGTGTTCCTTA 60
QY 61 TATAGATGGAACAAAGCGACGCAAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120
DB 61 TATAGATGGAACAAAGCGACGCAAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120
QY 121 GCGTGAAGGAATCAATGCTGAGCAATTCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180

AAV27351
 ID AAV27351 standard; DNA; 2389 BP.
 XX
 AC AAV27351;
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae SP0036 nucleotide.
 XX
 DE Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
 XX
 KW detection; pneumonia; otitis media; meningitis; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 PH Key Location/Qualifiers
 FT 2.2389
 FT CDS
 FT /*tag= a
 FT /product= "SP0036"
 FT /transl_except= (pos:1367..1369,aa:Xaa)
 FT /note= "no stop codon given; Xaa is unspecified"
 XX
 PN WO9818930-A2.
 XX
 XX 07-MAY-1998.
 XX
 XX 30-OCT-1997; 97WO-US019422.
 XX
 XX 31-OCT-1996; 96US-0029960P.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
 PI
 XX
 XX WPI; 1998-272224/24.
 DR P-PSDB; AAW55090.
 XX
 XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
 PT - or their epitope-containing fragments, useful in protective or
 PT therapeutic vaccines, and for diagnosis.
 XX
 XX Claim 1; Page 59; 118pp; English.
 PS
 XX
 CC The present sequence encodes a protein from Streptococcus pneumoniae. The
 CC nucleic acid sequence encoding the Streptococcus pneumoniae protein can
 CC be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose
 XX
 SQ Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;
 Query Match 100.0%; Score 2389; DB 2; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTCTTAGGCTGGGACTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA 60
 DB 1 TTCTTAGGCTGGGACTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA 60
 QY 61 TATAGATGGAAGAAACAAAGCGACGCAAAACCGAGAAATTCGACTCCTGATGAGGTAGCAA 120
 DB 61 TATAGATGGAAGAAACAAAGCGACGCAAAACCGAGAAATTCGACTCCTGATGAGGTAGCAA 120
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTAC 180

DB 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTAC 180
 QY 181 TTCATATGGCGACCACTATCATTTATTACATGGTAAGGTTTCCTTTATGACGCTATCATCAG 240
 DB 181 TTCATATGGCGACCACTATCATTTATTACATGGTAAGGTTTCCTTTATGACGCTATCATCAG 240
 QY 241 TGAAGAATTACTCATGAAGATCCAACTATAAGCTAAAGATGAGGATATTTGTTAATGA 300
 DB 241 TGAAGAATTACTCATGAAGATCCAACTATAAGCTAAAGATGAGGATATTTGTTAATGA 300
 QY 301 GGTCAAGGGTGGGATATGTTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360
 DB 301 GGTCAAGGGTGGGATATGTTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360
 QY 361 TGCCCAAGGGATTAACGTCGGTACAAAGAGGAAATCAATTCGACAAACAAAGAGCATAG 420
 DB 361 TGCCCAAGGGATTAACGTCGGTACAAAGAGGAAATCAATTCGACAAACAAAGAGCATAG 420
 QY 421 TCACATCGTGAAGGTGGAACTCCAAAGAACGATGGTGTGCTTGGCTTGGCAGCTTTCGCA 480
 DB 421 TCACATCGTGAAGGTGGAACTCCAAAGAACGATGGTGTGCTTGGCTTGGCAGCTTTCGCA 480
 QY 481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 DB 481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 QY 541 TGGTATGCTTATATCGTTCCTCATGAGATCAATTAACCATTAATTCCTTAAGAATGAGTT 600
 DB 541 TGGTATGCTTATATCGTTCCTCATGAGATCAATTAACCATTAATTCCTTAAGAATGAGTT 600
 QY 601 ATCAGCTAGCGAGTGGCTGCTGCAGAACGCTTCTATCTGGTGGAGGAAATCTGCAAA 660
 DB 601 ATCAGCTAGCGAGTGGCTGCTGCAGAACGCTTCTATCTGGTGGAGGAAATCTGCAAA 660
 QY 661 TTCAAGAACCTATCGCCGACAAATAAGCGATAACACTTCAAGAACAAATCTGGGTACTTTC 720
 DB 661 TTCAAGAACCTATCGCCGACAAATAAGCGATAACACTTCAAGAACAAATCTGGGTACTTTC 720
 QY 721 TGTAGCAATCCAGGAATCTACAAATACTAACCAAGCAACAGCAACAACTAACAGTCA 780
 DB 721 TGTAGCAATCCAGGAATCTACAAATACTAACCAAGCAACAGCAACAACTAACAGTCA 780
 QY 781 AGCAAGTCAAGTAAATGACATGATAGTCTTCTGAAACAGCTCTCAAAATCGCTTTGAG 840
 DB 781 AGCAAGTCAAGTAAATGACATGATAGTCTTCTGAAACAGCTCTCAAAATCGCTTTGAG 840
 QY 841 TCAACGACATGTAGAAATCTGATGGCCCTTGTCTTGATCCAGCACAATCAAGTCAAC 900
 DB 841 TCAACGACATGTAGAAATCTGATGGCCCTTGTCTTGATCCAGCACAATCAAGTCAAC 900
 QY 901 AGCTAGAGTGTGAGTGGCAGCGAGATCATACCACTTCCATCCCTTACTCTCAAAAT 960
 DB 901 AGCTAGAGTGTGAGTGGCAGCGAGATCATACCACTTCCATCCCTTACTCTCAAAAT 960
 QY 961 GTCTGAATTTGGAAGAACGAATCGCTGATATTTCCCTTCTGTTATCGTTTCAAAACATTG 1020
 DB 961 GTCTGAATTTGGAAGAACGAATCGCTGATATTTCCCTTCTGTTATCGTTTCAAAACATTG 1020
 QY 1021 GSTACAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG 1080
 DB 1021 GSTACAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG 1080
 QY 1081 CCGGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGT 1140
 DB 1081 CCGGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGT 1140
 QY 1141 ACMAAAGTTGGGAGGATATGTTTTCGAAGAAAGGCGCATCTCTCGTTATGCTTTTCG 1200
 DB 1141 ACMAAAGTTGGGAGGATATGTTTTCGAAGAAAGGCGCATCTCTCGTTATGCTTTTCG 1200
 QY 1201 GAAAGATTTACCATCTGAAACTGTTAAATAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260
 DB 1201 GAAAGATTTACCATCTGAAACTGTTAAATAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 30, 2004, 20:07:08 ; Search time 915 Seconds
(without alignments)
11091.743 Million cell updates/sec

Title: US-09-765-271-55
Perfect score: 2389
Sequence: 1 TTCTTACGAGTTGGAGCTG.....TAAGTAAGCAAAATAAATAAC 2389

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : N Geneseq 23Jan04 : *
1: geneseqn1980s : *
2: geneseqn1990s : *
3: geneseqn2000s : *
4: geneseqn2001s : *
5: geneseqn2001bs : *
6: geneseqn2002s : *
7: geneseqn2003s : *
8: geneseqn2003bs : *
9: geneseqn2003cs : *
10: geneseqn2004s : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | ID | Description |
|------------|-------|-------------|------------|---------------------|
| 1 | 2389 | 100.0 | 2 AAV27351 | AAV27351 Streptococ |
| 2 | 2389 | 100.0 | 6 ABQ84819 | Abq84819 S. pneumo |
| 3 | 2389 | 100.0 | 9 ADC45136 | Adc45136 S. pneumo |
| 4 | 2389 | 100.0 | 3 AAA47604 | AAA47604 Recombina |
| 5 | 2338 | 97.9 | 7 ABX06886 | Abx06886 S. pneumo |
| 6 | 2338 | 97.9 | 7 AAV52227 | AAV52227 Streptoco |
| 7 | 2338 | 97.9 | 8195 2 | Continuation (10 o |
| 8 | 2338 | 97.9 | 110000 7 | Continuation (11 o |
| 9 | 148 | 6.2 | 2523 3 | AAA65731 |
| 10 | 148 | 6.2 | 2647 3 | AAA65736 |
| 11 | 148 | 6.2 | 2647 6 | ABK15103 |
| 12 | 53 | 2.2 | 504 3 | AAA08856 |
| 13 | 53 | 2.2 | 2163 2 | AAV25394 |
| 14 | 53 | 2.2 | 2290 6 | ABQ84824 |
| 15 | 53 | 2.2 | 2290 6 | ABQ84824 S. pneumo |
| 16 | 53 | 2.2 | 2290 9 | ADC45146 |
| 17 | 53 | 2.2 | 2359 2 | AAV52376 |
| 18 | 53 | 2.2 | 2457 7 | ABX06885 |
| 19 | 53 | 2.2 | 2478 3 | AAA08557 |
| 20 | 53 | 2.2 | 2481 3 | AAA05417 |
| 21 | 53 | 2.2 | 2517 7 | ABX06705 |
| 22 | 53 | 2.2 | 2531 3 | AAA47605 |
| 23 | 53 | 2.2 | 2531 3 | AAA47602 |

ALIGNMENTS

RESULT 1

| | | | | | | |
|----|----|-----|--------|---|-------------|---------------------|
| 24 | 53 | 2.2 | 2639 | 3 | AAA65737 | AAA65737 Streptoco |
| 25 | 53 | 2.2 | 2639 | 7 | ABK15104 | Abk15104 DNA encod |
| 26 | 53 | 2.2 | 110000 | 6 | ABSS6454_08 | Continuation (9 of |
| 27 | 50 | 2.1 | 492 | 2 | AAV25393 | Aav25393 Streptoco |
| 28 | 44 | 1.8 | 973 | 2 | AAV52488 | Aav52488 Streptoco |
| 29 | 38 | 1.6 | 3171 | 3 | AAA65739 | AAA65739 Streptoco |
| 30 | 30 | 1.3 | 40 | 2 | AAV27492 | Aav27492 Streptoco |
| 31 | 30 | 1.3 | 40 | 2 | AAA47601 | AAA47601 Primer fo |
| 32 | 30 | 1.3 | 40 | 6 | ABQ84960 | Abq84960 Streptoco |
| 33 | 30 | 1.3 | 40 | 9 | ADC45363 | Adc45363 S. pneumo |
| 34 | 27 | 1.1 | 36 | 3 | AAA47599 | AAA47599 Primer fo |
| 35 | 27 | 1.1 | 37 | 2 | AAV27491 | Aav27491 Streptoco |
| 36 | 27 | 1.1 | 37 | 6 | ABQ84959 | Abq84959 Streptoco |
| 37 | 27 | 1.1 | 37 | 9 | ADC45362 | Adc45362 S. pneumo |
| 38 | 26 | 1.1 | 35 | 3 | AAA47600 | AAA47600 Primer fo |
| 39 | 25 | 1.0 | 33 | 3 | AAA65762 | AAA65762 Streptoco |
| 40 | 25 | 1.0 | 33 | 6 | ABK33684 | Abk33684 S. pneumo |
| 41 | 25 | 1.0 | 34 | 3 | AAA65764 | AAA65764 Streptoco |
| 42 | 25 | 1.0 | 34 | 6 | ABK33686 | Abk33686 S. pneumo |
| 43 | 25 | 1.0 | 35 | 6 | ABK33719 | Abk33719 S. pneumo |
| 44 | 24 | 1.0 | 1342 | 2 | AAV27414 | Aav27414 Streptoco |
| 45 | 24 | 1.0 | 1342 | 6 | ABQ84882 | Abq84882 S. pneumo |
| 46 | 24 | 1.0 | 1342 | 9 | ADC45262 | Adc45262 S. pneumo |
| 47 | 24 | 1.0 | 1398 | 6 | ABN66839 | ABN66839 Streptoco |
| 48 | 24 | 1.0 | 1455 | 3 | AAA05473 | AAA05473 Streptoco |
| 49 | 24 | 1.0 | 1455 | 3 | AAZ91804 | Aaz91804 Streptoco |
| 50 | 24 | 1.0 | 1455 | 3 | AAA65733 | AAA65733 Streptoco |
| 51 | 24 | 1.0 | 1455 | 3 | AAA47603 | AAA47603 Recombina |
| 52 | 24 | 1.0 | 2528 | 3 | AAA65738 | AAA65738 Streptoco |
| 53 | 24 | 1.0 | 2528 | 6 | ABK15105 | Abk15105 DNA encod |
| 54 | 24 | 1.0 | 3117 | 7 | ABX06706 | ABX06706 S. pneumo |
| 55 | 24 | 1.0 | 3120 | 3 | AAA65730 | AAA65730 Streptoco |
| 56 | 24 | 1.0 | 5048 | 3 | AAA65735 | AAA65735 Streptoco |
| 57 | 24 | 1.0 | 5048 | 6 | ABK15101 | Abk15101 DNA encod |
| 58 | 24 | 1.0 | 6867 | 2 | AAV52325 | AAV52325 Streptoco |
| 59 | 24 | 1.0 | 110000 | 6 | ABN71527_18 | Continuation (19 o |
| 60 | 23 | 1.0 | 77287 | 8 | AAD58261 | Aad58261 Murine tu |
| 61 | 23 | 1.0 | 167739 | 8 | AAD58258 | Aad58258 Murine tu |
| 62 | 22 | 0.9 | 1547 | 3 | AAK37581 | Aac37581 Arabidops |
| 63 | 21 | 0.9 | 29 | 3 | AAA65761 | AAA65761 Streptoco |
| 64 | 21 | 0.9 | 29 | 6 | ABK33683 | Abk33683 S. pneumo |
| 65 | 21 | 0.9 | 690 | 8 | ACL22740 | ACL22740 DNA clone |
| 66 | 21 | 0.9 | 110000 | 7 | ABQ84281_2 | Continuation (3 of |
| 67 | 20 | 0.8 | 501 | 6 | AAI69344 | AAI69344 C. albica |
| 68 | 20 | 0.8 | 1146 | 6 | AAA05814 | AAA05814 Group B S |
| 69 | 20 | 0.8 | 2466 | 6 | ABN69535 | ABN69535 Streptoco |
| 70 | 20 | 0.8 | 2466 | 6 | ABN70334 | ABN70334 Streptoco |
| 71 | 20 | 0.8 | 2469 | 3 | AAA05811 | AAA05811 Group B S |
| 72 | 20 | 0.8 | 2469 | 3 | AAA65740 | AAA65740 Streptoco |
| 73 | 20 | 0.8 | 2469 | 5 | AAA00038 | AAA00038 Streptoco |
| 74 | 20 | 0.8 | 2472 | 6 | AAA65741 | AAA65741 Streptoco |
| 75 | 20 | 0.8 | 2478 | 6 | ABN66838 | ABN66838 Streptoco |
| 76 | 20 | 0.8 | 2478 | 5 | AAA00036 | AAA00036 Streptoco |
| 77 | 20 | 0.8 | 5158 | 6 | AAI69348 | AAI69348 C. albica |
| 78 | 20 | 0.8 | 5215 | 6 | AAIX91105 | AAIX91105 Group B S |
| 79 | 20 | 0.8 | 110000 | 6 | ABN71527_12 | Continuation (13 o |
| 80 | 19 | 0.8 | 164 | 7 | ABX54820 | ABX54820 Bovine ES |
| 81 | 19 | 0.8 | 426 | 5 | ABV02215 | ABV02215 Human pro |
| 82 | 19 | 0.8 | 427 | 4 | AAK30789 | AAK30789 Human CDN |
| 83 | 19 | 0.8 | 468 | 5 | ABV11384 | ABV11384 Human pro |
| 84 | 19 | 0.8 | 481 | 5 | ABV32529 | ABV32529 Human pro |
| 85 | 19 | 0.8 | 481 | 5 | ABV41454 | ABV41454 Human pro |
| 86 | 19 | 0.8 | 693 | 6 | ABQ89355 | ABQ89355 Human pro |
| 87 | 19 | 0.8 | 693 | 8 | ADB82298 | ADB82298 Human CDN |
| 88 | 19 | 0.8 | 1020 | 7 | ACA28893 | ACA28893 Prokaryot |
| 89 | 19 | 0.8 | 1203 | 5 | ABV21401 | ABV21401 Human pro |
| 90 | 19 | 0.8 | 1205 | 5 | ABV27219 | ABV27219 Human pro |

| | Query Match | 2.2% | Score 53 | DB 6 | Length 2230 | |
|----|-----------------------|---------------|-------------------|---------------------|-------------|--------|
| | Best Local Similarity | 100.0% | Pred. No. 2.2e-16 | | | |
| | Matches 53 | Conservative | 0 | Mismatches 0 | Indels 0 | Gaps 0 |
| Qy | 983 | GCTCGTATTATTC | CCCTTCGTTATCGTT | CAACCAACCATTCGGGTAC | CAGATTCAAG | 1035 |
| Dp | 944 | GTCGTATTATTC | CCCTTCGTTATCGTT | CAACCAACCATTCGGGTAC | CAGATTCAAG | 996 |

KEYWORDS JP 2002526082-A/1.
SOURCE Streptococcus pneumoniae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 504)
AUTHORS Hostetter, M.K., Finkel, D.J., Cheng, Q., Green, B.A. and Masi, A.W.
TITLE Human complement C3-degrading polypeptides from streptococcus
JOURNAL Patent: JP 2002526082-A 1 20-AUG-2002;
REGENTS OF THE UNIVERSITY OF MINNESOTA, AMERICAN CYANAMID CO
COMMENT OS Streptococcus pneumoniae
PN JP 2002526082-A/1
PD 20-AUG-2002
PF 24-SEP-1999 JP 2000574269
PR 24-SEP-1998 US 60/101736, 31-MAR-1999 US 09/283094 PI
MARGARET K HOSTETTER, DAVID J FINKEL, QI CHENG, BRUCE A GREEN, AMY
PI W NASI
PC C12N15/09, A61K38/00, A61K39/09, A61K39/395, A61K48/00,
A61P29/00,
PC A61P37/04, A61P37/06, A61P43/00, C07K16/40, C12N1/15, C12N1/19, PC
C12N1/21,
PC C12N5/10, C12N9/48, C12P21/08, C12N15/00, C12N5/00, A61K37/02 CC
Human complement C3-degrading polypeptides from streptococcus CC
pneumoniae

FH Key Location/Qualifiers
FT source 1..504
FT /organism='Streptococcus pneumoniae'.
/mol_type='genomic DNA'
/db_xref='taxon:1313'

ORIGIN
Query Match 2.2%; Score 53; DB 6; Length 504;
Best Local Similarity 100.0%; Pred. No. 2.4e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GTCGCTATTATTCCTTCGTTATCGTTCAACACCATGGGTACGATTCAG 1035
|||||
Db 184 GTCGCTATTATTCCTTCGTTATCGTTCAACACCATGGGTACGATTCAG 236
|||||

RESULT 21
BD229971/c
LOCUS
DEFINITION Human complement C3-degrading polypeptides from streptococcus pneumoniae.
ACCESSION BD229971.1 GI:33039741
VERSION BD229971.1
KEYWORDS JP 2002526082-A/2.
SOURCE Streptococcus pneumoniae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 504)
AUTHORS Hostetter, M.K., Finkel, D.J., Cheng, Q., Green, B.A. and Masi, A.W.
TITLE Human complement C3-degrading polypeptides from streptococcus
JOURNAL Patent: JP 2002526082-A 2 20-AUG-2002;
REGENTS OF THE UNIVERSITY OF MINNESOTA, AMERICAN CYANAMID CO
COMMENT OS Streptococcus pneumoniae
PN JP 2002526082-A/2
PD 20-AUG-2002
PF 24-SEP-1999 JP 2000574269
PR 24-SEP-1998 US 60/101736, 31-MAR-1999 US 09/283094 PI
MARGARET K HOSTETTER, DAVID J FINKEL, QI CHENG, BRUCE A GREEN, AMY
PI W NASI
PC C12N15/09, A61K38/00, A61K39/09, A61K39/395, A61K48/00,
A61P29/00,
PC A61P37/04, A61P37/06, A61P43/00, C07K16/40, C12N1/15, C12N1/19, PC
C12N1/21,
PC C12N5/10, C12N9/48, C12P21/08, C12N15/00, C12N5/00, A61K37/02 CC
Human complement C3-degrading polypeptides from streptococcus CC

pneumoniae
FH Key Location/Qualifiers
FT source 1..504
FT /organism='Streptococcus pneumoniae'.
/mol_type='genomic DNA'
/db_xref='taxon:1313'

FEATURES
source
1..504
/organism='Streptococcus pneumoniae'
/mol_type='genomic DNA'
/db_xref='taxon:1313'

ORIGIN
Query Match 2.2%; Score 53; DB 6; Length 504;
Best Local Similarity 100.0%; Pred. No. 2.4e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GTCGCTATTATTCCTTCGTTATCGTTCAACACCATGGGTACGATTCAG 1035
|||||
Db 321 GTCGCTATTATTCCTTCGTTATCGTTCAACACCATGGGTACGATTCAG 269
|||||

RESULT 22
AF340222
LOCUS
DEFINITION Synthetic construct PhpA-79 (phpA-79) gene, complete cds.
ACCESSION AF340222
VERSION AF340222.1 GI:13447095
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 2166)
AUTHORS Zhang, Y., Masi, A.W., Barniak, V., Mountzouros, K., Hostetter, M.K. and Green, B.A.
TITLE Recombinant PhpA protein, a unique histidine motif-containing protein from Streptococcus pneumoniae, protects mice against intranasal pneumococcal challenge
JOURNAL Infect. Immun. 69 (6), 3827-3836 (2001)
MEDLINE 21246685
PubMed 11349048
REFERENCE 2 (bases 1 to 2166)
AUTHORS Zhang, Y., Masi, A., Barniak, V., Mountzouros, K., Hostetter, M. and Green, B.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2001) Department of Bacteriology, Wyeth Lederle Vaccines, 211 Bailey Road, West Henrietta, NY 14586, USA
FEATURES
source
1..2166
/organism='synthetic construct'
/mol_type='genomic DNA'
/db_xref='taxon:32630'
/note='derived from Streptococcus pneumoniae PhpA sequence'
1..2166
/gene='phpA-79'
1..2166
/gene='phpA-79'
/note='N-terminal truncated PhpA protein; histidine motif-containing protein'
/codon_start=1
/transl_table=11
/product='PhpA-79'
/protein_id='AAK26630.1'
/db_xref='GI:13447096'
/translation='MKDQNYQLKSDIWEIKGVIVKDGKYYVYLKDAADNRITKEEIKRQKHSHNGGSDQAVVAARAGQRYTTDDGYIFNADSIIEDTGDAYIVPHGDHYIIPKNEISASELAEEAYWNGKQSRPSSSSSYNANPAQRLSENHNLVTPTYHONGENISLLRELKPLSRHVESDGLIFDPAQITSRARGVAVPHGNHVFIPYBOMSELKRTARIIPLYVRNHWVPDSEPEPSPQSTPEPSPQAPNPPASNPIDELKLVKAVRKVGCVGFEENGVSRYIPAKDLSAETAGIDSKLAKGSLKUGAKKIDLPSSDRFYNKAYDLIARLHQDLNKGQVDFEALNLLKLDKVPDSKVLVDLILAFAPRHPERLGNPAQITTYDDEIQVALAGKTYTDEGYIFDRPITSDRGA VYPTWHTSHWIKDLSAEARAAQAYAKEKGLTPPSTDHODSNTAKAGAEAIYNRVKAACKVPLDRMPYNLQYTVKNGSLIIPHYDHYHNKFEWFDSEGLVEAPKGYTLED

```

RESULT 17
AX343072          AX343072          2523 bp          PAT 12-JAN-2002
LOCUS              Sequence 3 from Patent WO0198334.
DEFINITION          Streptococcus antigens
ACCESSION            AX343072
VERSION              AX343072.1 GI:18152270
KEYWORDS              unidentified
SOURCE                unidentified
ORGANISM              unclassified.
REFERENCE
1
AUTHORS              Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
TITLE                Streptococcus antigens
JOURNAL              Patent: WO 0198334-A 3 27-DEC-2001;
                     SHIRE BIOCHEM INC. (CA)
FEATURES
source
1..2523
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
/note="BVH-11"
ORIGIN

Query Match          6.2%; Score 148; DB 6; Length 2523;
Best Local Similarity 100.0%; Pred. No. 1.1e-67;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 699 CAAGAACAACCTGGGTACCTTCTGTAGCAATCCAGGAACCTACAATACTAACACAAGCA 758
Db 758 CAAGAACAACCTGGGTACCTTCTGTAGCAATCCAGGAACCTACAATACTAACACAAGCA 817
Qy 759 ACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAAC 818
Db 818 ACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAAC 877
Qy 819 AGCTCTACAACTGCCTTTGAGTCAACG 846
Db 878 AGCTCTACAACTGCCTTTGAGTCAACG 905

RESULT 18
BD263590          BD263590          2647 bp          PAT 17-JUL-2003
LOCUS              Novel streptococcus antigens.
DEFINITION          Streptococcus antigens.
ACCESSION            BD263590
VERSION              BD263590.1 GI:33073358
KEYWORDS              unidentified
SOURCE                unidentified
ORGANISM              unclassified.
REFERENCE
1 (bases 1 to 2647)
AUTHORS              Hamel,J., Brodeur,B.R., Pineau,I., Martin,D., Rioux,C. and
                     Charland,N.
TITLE                Novel streptococcus antigens
JOURNAL              Patent: JP 2002533123-A 7 08-OCT-2002;
                     SHIRE BIOCHEM INC
COMMENT              OS S. pneumoniae
                     PN JP 2002533123-A/7
                     PD 08-OCT-2002
                     PF 20-DEC-1999 JP 2000591190
                     PR 23-DEC-1998 US 60/113800
                     PI JOSEF HAMEL, BERNARD R BRODEUR, ISABELLE PINEAU, DENIS MARTIN, PI
                     CLEMENT RIOUX,
                     PI NATHALIE CHARLAND
                     PC C12N15/09,A61K39/00,A61K39/39,A61P11/00,A61P25/00,
                     PC A61P27/16,
                     PC A61P31/04,C07K14/315,C07K19/00,C12N1/15,C12N1/19,C12N1/21, PC
                     C12N5/10,
                     PC C12P21/02,C12N15/00,C12N5/00
                     CC Novel streptococcus antigens
                     FH Key
                     Location/Qualifiers

```

```

FT source          1..2647
FT                 /organism='S. pneumoniae'.
FEATURES
source
1..2647
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN

Query Match          6.2%; Score 148; DB 6; Length 2647;
Best Local Similarity 100.0%; Pred. No. 1.1e-67;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 699 CAAGAACAACCTGGGTACCTTCTGTAGCAATCCAGGAACCTACAATACTAACACAAGCA 758
Db 802 CAAGAACAACCTGGGTACCTTCTGTAGCAATCCAGGAACCTACAATACTAACACAAGCA 861
Qy 759 ACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAAC 818
Db 862 ACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAAC 921
Qy 819 AGCTCTACAACTGCCTTTGAGTCAACG 846
Db 922 AGCTCTACAACTGCCTTTGAGTCAACG 949

RESULT 19
AX343073          AX343073          2647 bp          PAT 12-JAN-2002
LOCUS              Sequence 4 from Patent WO0198334.
DEFINITION          Streptococcus antigens
ACCESSION            AX343073
VERSION              AX343073.1 GI:18152271
KEYWORDS              unidentified
SOURCE                unidentified
ORGANISM              unclassified.
REFERENCE
1
AUTHORS              Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
TITLE                Streptococcus antigens
JOURNAL              Patent: WO 0198334-A 4 27-DEC-2001;
                     SHIRE BIOCHEM INC. (CA)
FEATURES
source
1..2647
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
/note="BVH-11"
ORIGIN

Query Match          6.2%; Score 148; DB 6; Length 2647;
Best Local Similarity 100.0%; Pred. No. 1.1e-67;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 699 CAAGAACAACCTGGGTACCTTCTGTAGCAATCCAGGAACCTACAATACTAACACAAGCA 758
Db 802 CAAGAACAACCTGGGTACCTTCTGTAGCAATCCAGGAACCTACAATACTAACACAAGCA 861
Qy 759 ACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAAC 818
Db 862 ACAACAGCAACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAAC 921
Qy 819 AGCTCTACAACTGCCTTTGAGTCAACG 846
Db 922 AGCTCTACAACTGCCTTTGAGTCAACG 949

RESULT 20
BD229970          BD229970          504 bp          PAT 17-JUL-2003
LOCUS              Human complement C3-degrading polypeptides from streptococcus
DEFINITION          pneumoniae.
ACCESSION            BD229970
VERSION              BD229970.1 GI:33039740

```

```

Qy 686 AGCGATACACTTCAAGAACAACTGGGTACCTTCTGTGAAGCAATCCAGGAACCTACAAT 745
Db 468 AGCGATACACTTCAAGAACAACTGGGTACCTTCTGTGAAGCAATCCAGGAACCTACAAT 527
Qy 746 ACTAACACAGCAACACAGCAACACTAACAGTCAAGCAAGTCAAGGTAATGACATTGAT 805
Db 528 ACTAACACAGCAACACAGCAACACTAACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 587
Qy 806 AGTCTCTTGAACAGCTCTCAAACTGCTTGTGAGTCAACG 846
Db 588 AGTCTCTTGAACAGCTCTCAAACTGCTTGTGAGTCAACG 628

RESULT 15
AF340221 2535 bp DNA linear BCT 22-MAY-2001
LOCUS Streptococcus pneumoniae PnpA (pnpA) gene, complete cds.
ACCESSION AF340221
VERSION AF340221.1 GI:13447093
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 2535)
Zhang, Y., Masi, A.W., Barniak, V., Mountzouros, K., Hostetter, M.K. and
Green, B.A.
Recombinant PnpA protein, a unique histidine motif-containing
protein from Streptococcus pneumoniae, protects mice against
intranasal pneumococcal challenge
Infect. Immun. 69 (6), 3827-3836 (2001)
21246685
PUBMED 11349048
REFERENCE 2 (bases 1 to 2535)
Zhang, Y., Masi, A., Barniak, V., Mountzouros, K., Hostetter, M. and
Green, B.
Direct Submission
TITLE Streptococcus pneumoniae
JOURNAL Submitted (25-JAN-2001) Department of Bacteriology, Wyeth Lederle
Vaccines, 211 Bailey Road, West Henrietta, NY 14586, USA
FEATURES
source
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"
1..2535
/gene="pnpA"
1..2535
/gene="pnpA"
/notes="histidine motif-containing protein"
/codon_start=1
/transl_table=11
/product="pnpA"
/protein_id="AAK26629.1"
/db_xref="GI:13447094"
/translation="MKINKLYVGSAAALILVCSYELGLYQARTVKENRVSIDGK
QATKTENJTPDVSREGINAGIIVIKITDQGVTSHGHDHYHYNGKVPYDAIISEE
LMKDPNYOLKDEDIISIKGVYIVKGVYVYLKDAHADNVRTKEINRQKQHS
QHQEGGTPRNDGVALARSQVTTDDGVIFNADSIIEDTGDAIVPHGDHYHVPKN
ELASLAAKAFILSGRGLNSRTYRQNSDTSRTNVPVSVNPGTNTINISNSN
TNSAQSDNDISLLKQLYKLPUSQRHVSDELGFDPQITSTANGVAVPHGDHYHF
IPYSQSPUEKLARIIPURYSNHWVPDSRPEQSPQSTPEPSQPAFPNPQAPS
NIDEKLKVEARFNGVGVYFEENGVPYIPAKDLAETAAAGIDSKLAKQESLSHLG
AKTDLFSSDRFENKAYDLARIHQDLNDKGRQVDFAALDNLRLERLKVSDVKVL
VDDILAFAPLRPERLKPNAQITVTDDEIOVAKLAGKYTDEDGVIIPDRDITSEG
DAVTPHMTSHWIKDSISEARAAQAYAKEGLTPSTDHQDSGNTFEAKGAAYI
NRVKAQKVLDRMPNLQITVEVKNGLIIPHYDHYNIKEFWDEGLYEAPKGYSL
EDLATVYVVEHPNRPNSHDNGFNASDHVRNKNQADNTQTEKNEKPKQTEKE
EETPREKQSEKPEKTEPEESPESEPOVETEKVKEKLEAREADLLGK
IQNPIIKSNAKETLTGLKNLLFGTQDNNTIMAEAEKLLALLKESK"

Qy 819 AGCTCTACAACTGCGCTTTGAGTCAACG 846
Db 878 AGCTCTACAACTGCGCTTTGAGTCAACG 905

Query Match 8.4%; Score 200; DB 1; Length 2535;
Best Local Similarity 100.0%; Pred. No. 9.3e-96;

```

```

Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 647 GGAAATCTGTCAAATTCAGAACCTATCGCGGCAAAATAGCGATAACACTTCAAGAACA 706
Db 706 GGAAATCTGTCAAATTCAGAACCTATCGCGGCAAAATAGCGATAACACTTCAAGAACA 765
Qy 707 AACTGGGTACCTTCTGTGAAGCAATCCAGGAACCTACAATACTACAAGCAACAGC 766
Db 766 AACTGGGTACCTTCTGTGAAGCAATCCAGGAACCTACAATACTACAAGCAACAGC 825
Qy 767 AACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAACAGCTCTAC 826
Db 826 AACACTAACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAACAGCTCTAC 885
Qy 827 AACTGCGCTTTGAGTCAACG 846
Db 886 AACTGCGCTTTGAGTCAACG 905

RESULT 16
BD263585 2523 bp DNA linear PAT 17-JUL-2003
LOCUS Novel streptococcus antigens.
ACCESSION BD263585
VERSION BD263585.1 GI:33073353
KEYWORDS JP 2002533123-A/2.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 2523)
Hamel, J., Brodeur, B.R., Pineau, I., Martin, D., Rioux, C. and
Charland, N.
Novel streptococcus antigens
Patent: JP 2002533123-A 2 08-OCT-2002;
SHIRE BIOCHEM INC
COMMENT
OS S. pneumoniae
PN JP 2002533123-A/2
PD 08-OCT-2002
PF 20-DEC-1999 JP 2000591190
PR 23-DEC-1998 US 60/113800
PI JOSEEE HAMEL, BERNARD R BRODEUR, ISABELLE PINEAU, DENIS MARTIN, PI
CLEMMENT RIOUX,
PI NATHALIE CHARLAND
PC C12N15/02, C12N15/00, C12N5/00
PC A61P27/16,
PC A61P31/04, C07K14/315, C07K19/00, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10,
PC C12P21/02, C12N15/00, C12N5/00
CC Coding region of BVH-11 gene
FH Key Location/Qualifiers
FT CDS (1)..(2520).
FEATURES
source
1..2523
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Query Match 6.2%; Score 148; DB 6; Length 2523;
Best Local Similarity 100.0%; Pred. No. 1.1e-67;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 699 CAAGACAACTGGGTACCTTCTGTGAAGCAATCCAGGAACCTACAATACTACAAGCA 758
Db 758 CAAGACAACTGGGTACCTTCTGTGAAGCAATCCAGGAACCTACAATACTACAAGCA 817
Qy 759 ACAACAGCAACTACTACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAAC 818
Db 818 ACAACAGCAACTACTACAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTTGAAC 877
Qy 819 AGCTCTACAACTGCGCTTTGAGTCAACG 846
Db 878 AGCTCTACAACTGCGCTTTGAGTCAACG 905

```

Db 18433 TTTGAAATAAGGTCGTAATCTGATTTCAASCCTTGGCAAAATTTTGAAGCGCTTG 18492
 QY 1424 AATGATGAATCGAC-TAATAAGAAAAATTTGGTAGATGATTTATTTGGCATTCCTAGCACC 1482
 Db 18493 AATGATGAATCGACTTAATAAGAAAAATTTGGTAGATGATTTATTTGGCATTCCTAGCACC 18552
 QY 1483 AATTACCACCGAGCGACTTGGCAACCAAAATTTCTAAATTGAGTATACCTAGAGAGA 1542
 Db 18553 AATTACCACCGAGCGACTTGGCAACCAAAATTTCTAAATTGAGTATACCTAGAGAGA 18612
 QY 1543 AGTTGCTATTGCTCAATTTAGCTGTAAGTATACAACGTCAGATGGTTACATTTTGGATGA 1602
 Db 18613 AGTTGCTATTGCTCAATTTAGCTGTAAGTATACAACGTCAGATGGTTACATTTTGGATGA 18672
 QY 1603 ACATGATATAATCAGTGAAGAGATGCAATATGTAACGCTCATATGGGCCATAGTCA 1662
 Db 18673 ACATGATATAATCAGTGAAGAGATGCAATATGTAACGCTCATATGGGCCATAGTCA 18732
 QY 1663 CTGGATTGGAAAAGATAGCTTTCTGATAAGAAAAAGTTGCAGCTCAGACCTATACTAA 1722
 Db 18733 CTGGATTGGAAAAGATAGCTTTCTGATAAGAAAAAGTTGCAGCTCAGACCTATACTAA 18792
 QY 1723 AGAAAAAGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGA 1782
 Db 18793 AGAAAAAGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGA 18852
 QY 1783 TAGTCAGCAGCATTTTACAATCGTGTGAAGGGGAAAAAGAAATTCCTGTCGACT 1842
 Db 18853 TAGTCAGCAGCATTTTACAATCGTGTGAAGGGGAAAAAGAAATTCCTGTCGACT 18912
 QY 1843 TCCATATATGTTGAGCATACAGTTGAGGTTAAAGCAAGTAAATTTGATTTTCCATCAA 1902
 Db 18913 TCCATATATGTTGAGCATACAGTTGAGGTTAAAGCAAGTAAATTTGATTTTCCATCAA 18972
 QY 1903 GGATCATTTACCAATAATTTAAATTTGCTGTTGATGATCACATACAAAGCTCCAAA 1962
 Db 18973 GGATCATTTACCAATAATTTAAATTTGCTGTTGATGATCACATACAAAGCTCCAAA 19032
 QY 1963 TGGCTATACCTTGGAGATTTGTTGCGACGATTAAGTACTACGTAGAAACACCTGACGA 2022
 Db 19033 TGGCTATACCTTGGAGATTTGTTGCGACGATTAAGTACTACGTAGAAACACCTGACGA 19092
 QY 2023 AGTCCACATTTCAATGATGGATGGGCAATGCCAGTGAGCATGCTTGAAGCAAGAAAGA 2082
 Db 19093 AGTCCACATTTCAATGATGGATGGGCAATGCCAGTGAGCATGCTTGAAGCAAGAAAGA 19152
 QY 2083 CCACAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGCCAGTAGAGGAAACACC 2142
 Db 19153 CCACAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGCCAGTAGAGGAAACACC 19212
 QY 2143 TGTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAGAGAC 2202
 Db 19213 TGTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAGAGAC 19272
 QY 2203 AGAAGTTTTCCTTGGCAAGTACGCAATTCAGTCTGAAAGCCAAATGCAACGAACTCT 2262
 Db 19273 AGAAGTTTTCCTTGGCAAGTACGCAATTCAGTCTGAAAGCCAAATGCAACGAACTCT 19332
 QY 2263 AGCTGTTTACGAAATAATTTGACTCTTCAAATTTATGGATAACAATGATATCATGCGAGA 2322
 Db 19333 AGCTGTTTACGAAATAATTTGACTCTTCAAATTTATGGATAACAATGATATCATGCGAGA 19392
 QY 2323 AGCAGAAAAATTTACTTGCCTGTTTAAAGGAAGTAACTCTTCACTCTGTAAGTAAAGAAA 2382
 Db 19393 AGCAGAAAAATTTACTTGCCTGTTTAAAGGAAGTAACTCTTCACTCTGTAAGTAAAGAAA 19452
 QY 2383 AATAAAC 2389
 Db 19453 AATAAAC 19459

RESULT 14

SPNEU1929
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 source

SPNEU1929
 Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.
 AL449951
 HTG: HTGS PHASE2.
 Streptococcus pneumoniae
 Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 1
 Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Humbert, Y.,
 Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
 Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.,
 and Garcia-Bustos, J. F.
 Annotated draft genomic sequence from a Streptococcus pneumoniae
 type 19f clinical isolate
 Microb. Drug Resist. 7 (2), 99-125 (2001)
 21335329
 11442348
 2 (bases 1 to 702)
 Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
 Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
 Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and
 Garcia-Bustos, J. F.
 Direct Submission
 Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
 Severo Ochoa 2, 28760 Tres Cantos, SPAIN
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 Location/Qualifiers
 1..702
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /serotype="19f"
 /db_xref="taxon:1313"
 /clone="G54"

ORIGIN

Query Match 13.6%; Score 326; DB 2; Length 702;
 Best Local Similarity 99.1%; Pred. No. 9.7e-164;
 Matches 576; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 265 AACTATAAGCTAAAGATCAGGATATTGTTAATGAGTCAAGGTCGATATGTTATCAAG 325
 Db 48 AACTATAAGCTAAAGATCAGGATATTGTTAATGAGTCAAGGTCGATATGTTATCAAG 107
 QY 326 GTAGATGGAAAAATCTATCTTTACCTTAAGGATGCTGCCACCGGATAACGTCGGTACA 385
 Db 108 GTAGATGGAAAAATCTATCTTTACCTTAAGGATGCTGCCACCGGATAACGTCGGTACA 167
 QY 386 AAGAGGAAATCATCGACAAAAACAAGACATAGTCAACATCGTCAAGGTGAACCTCCA 445
 Db 168 AAGAGGAAATCAATCGACAAAAACAAGACATAGTCAACATCGTCAAGGTGAACCTCCA 227
 QY 445 AGAAACGATGGTCTGCTTGCCTTGGCACGTTCCGAAGGACGCTATATACTACAGATGATGT 505
 Db 228 AGAAACGATGGTCTGCTTGCCTTGGCACGTTCCGAAGGACGCTATATACTACAGATGATGT 287
 QY 506 TATATCTTTAATGCTTCTGATATCATAGAGGATATCGGTGATGCTTATATCGTTCCTCAT 565
 Db 288 TATATCTTTAATGCTTCTGATATCATAGAGGATATCGGTGATGCTTATATCGTTCCTCAT 347
 QY 566 GAGATCATTTACCATTTACATTCCTTAAGATGAGTTATCAGTACGAGTGGTGGTGTGCA 625
 Db 348 GAGATCATTTACCATTTACATTCCTTAAGATGAGTTATCAGTACGAGTGGTGGTGTGCA 407
 QY 626 GAAGCCTTCCTATCTGTCGAGGAAATCTCTCAAAATTCAGAAACCTTATCCCGCAAAAAT 685
 Db 408 GAAGCCTTCCTATCTGTCGAGGAAATCTCTCAAAATTCAGAAACCTTATCCCGCAAAAAT 467

Db 6428 TAATCTGATTTCCAAAGCCTTACACAAATTAATTAGAACGCTTGAATGATGAATCGACTAA 6369
 QY 1441 TAAAGAAAAATGGTAGATGATTATTTGGCAATTCCTAGCACCAATTTACCAATCCAGAGCG 1500
 Db 6368 TAAAGAAAAATGGTAGATGATTATTTGGCAATTCCTAGCACCAATTTACCAATCCAGAGCG 6309
 QY 1501 ACTTGGCAAAACCAATTTCTCAATTTGAGTATATCTGAAGACGAAGTTTCGTTATTTGCTCAAT 1560
 Db 6308 ACTTGGCAAGCCAAATTTCTCAATTTGAGTATATCTGAAGACGAAGTTTCGTTATTTGCTCAAT 6249
 QY 1561 ACCTGATAGTATACAACTGAGTGGTTACATTTTGTGTAACATGATATATATCAATCAAT 1620
 Db 6248 ACCTGATAGTATACAACTGAGTGGTTACATTTTGTGTAACATGATATATATCAATCAAT 6189
 QY 1621 TGAAGGAGATGATATGTAACCCCTCATATGGCCATAGTCACTGGATTGGAAAAAGATAG 1680
 Db 6188 TGAAGGAGATGATATGTAACCCCTCATATGGCCATAGTCACTGGATTGGAAAAAGATAG 6129
 QY 1681 CTTTCTGATAGGAAAAAGTTGCACTCAAGCCTATATCTAAAGAAAAAGGTATCTTACC 1740
 Db 6128 CTTTCTGATAGGAAAAAGTTGCACTCAAGCCTATATCTAAAGAAAAAGGTATCTTACC 6069
 QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATGATGACGAGCTTATTTA 1800
 Db 6068 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATGATGACGAGCTTATTTA 6009
 QY 1801 CAATCGTGTGAAGGGGAAAAAGCAATTCCTCGTTCGACTTCCATATATGTTGGTGAAGA 1860
 Db 6008 CAATCGTGTGAAGGGGAAAAAGCAATTCCTCGTTCGACTTCCATATATGTTGGTGAAGA 5949
 QY 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTCTCTCATAGGATCAATTAACCAATAT 1920
 Db 5948 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTCTCTCATAGGATCAATTAACCAATAT 5889
 QY 1921 TAAATTTGCTTGGTTGATGATACACATACAAAGCTCCAAATGGCTTATACCTTGGAGA 1980
 Db 5888 TAAATTTGCTTGGTTGATGATACACATACAAAGCTCCAAATGGCTTATACCTTGGAGA 5829
 QY 1981 TTTGTTTGGCAGATTAAGTACTAGTAGAACACCTGACCAAGCTCCACATCTCTAATGA 2040
 Db 5828 TTTGTTTGGCAGATTAAGTACTAGTAGAACACCTGACCAAGCTCCACATCTCTAATGA 5769
 QY 2041 TGGATGGGCAATGCCAGTGAAGATGTTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100
 Db 5768 TGGATGGGCAATGCCAGTGAAGATGTTTAGGCAAGAAAGACACAGTGAAGATCCAAA 5709
 QY 2101 TAAGAACTTCAAAGCGATGAAGACGAGTGAAGAAACACCTGCTGAGCAGAGTCCC 2160
 Db 5708 TAAGAACTTCAAAGCGATGAAGACGAGTGAAGAAACACCTGCTGAGCAGAGTCCC 5649
 QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAAGAACGAGAGTTTGTTCGGA 2220
 Db 5648 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAAGAACGAGAGTTTGTTCGGA 5589
 QY 2221 AGTAACGATTTCTAGTCTGAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
 Db 5588 AGTAACGATTTCTAGTCTGAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 5529
 QY 2281 TTTGACTCTTCAAATTTATGATAACATAGTATCATGCGAGAGCAGAAAAATCTTGC 2340
 Db 5528 TTTGACTCTTCAAATTTATGATAACATAGTATCATGCGAGAGCAGAAAAATCTTGC 5469
 QY 2341 GTTGTAAAGGAGTAATCTTCTCATCTGTAAGTAAAGGAAAAATAAAC 2389
 Db 5468 GTTGTAAAGGAGTAATCTTCTCATCTGTAAGTAAAGGAAAAATAAAC 5420

RESULT 13
 SPNEU1915
 LOCUS
 DEFINITION
 Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.
 AL449937
 VERSION
 AL449937.1 GI:11545162

KEYWORDS HTG; HTGS PHASE2.
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1
 AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
 Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
 Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
 and Garcia-Bustos,J.F.
 Annotated draft genomic sequence from a Streptococcus pneumoniae
 type 19F clinical isolate
 Microb. Drug Resist. 7 (2), 99-125 (2001)
 21335329
 11442348
 2 (bases 1 to 20035)
 Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
 Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
 Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
 Garcia-Bustos,J.F.
 Direct Submission
 Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
 Severo Ochoa 2, 28760 Tres Cantos, SPAIN
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 Location/Qualifiers
 1. 20035
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /serotype="19F"
 /db_xref="taxon:1313"
 /clone="G54"
 ORIGIN
 Query Match 38.8%; Score 927; DB 2; Length 20035;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 1497; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
 QY 884 CAAATCAAAAGTCGAACAGCTAGAGGTGTCAGTGCACACGAGATCAATACCATTTC 943
 Db 17953 CAAATCAAAAGTCGAACAGCTAGAGGTGTCAGTGCACACGAGATCAATACCATTTC 18012
 QY 944 ATCCCTTACTCTCAAAATGCTGAATTTGGAAGACGAATCGCTGATTTATTCCTTCGT 1003
 Db 18013 ATCCCTTACTCTCAAAATGCTGAATTTGGAAGACGAATCGCTGATTTATTCCTTCGT 18072
 QY 1004 TATCGTTCAACCAATTTGGGTACCAAGTTCAAGCCAGAACCAAGTCCACACCGACT 1063
 Db 18073 TATCGTTCAACCAATTTGGGTACCAAGTTCAAGCCAGAACCAAGTCCACACCGACT 18132
 QY 1064 CCGGAACCTAGTCCAGCCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATTCCTTCT 1123
 Db 18133 CCGGAACCTAGTCCAGCCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATTCCTTCT 18192
 QY 1124 TTGTTAGTCACTGTGTAAGAAAAGTTGGGAAGGATATGTTTCCGAAGAAAGGCATC 1183
 Db 18193 TTGTTAGTCACTGTGTAAGAAAAGTTGGGAAGGATATGTTTCCGAAGAAAGGCATC 18252
 QY 1184 TCTCGTTATGCTTTGCGAAAGTTTACCATCTGAAATCTTAAATCTTGAAGCAAG 1243
 Db 18253 TCTCGTTATGCTTTGCGAAAGTTTACCATCTGAAATCTTAAATCTTGAAGCAAG 18312
 QY 1244 TTATCAAAAACAGAGAGTCTTTTACACACTTTTAACTGCTTAAAGAAAGAAATCTGCTCT 1303
 Db 18313 TTATCAAAAACAGAGAGTCTTTTACACACTTTTAACTGCTTAAAGAAAGAAATCTGCTCT 18372
 QY 1304 CTGACCAAGAAATTTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGG 1363
 Db 18373 CTGACCAAGAAATTTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGG 18432
 QY 1364 TTGGAATAAGGGTCGTAATTTCTGATTTCCAGCCTTAGACAAATTTAGAACGCTTG 1423

| | | | | | | | | | | | |
|--|------|---|--------------|-------|----|------------|--------|--------|----|------|----|
| GVAVPHGDHYHFIPIYSOMSELEERIALIIPLYRSNHWVDSRPEQSPQTPPESPCQ POAPNLIKIDSNSSLSQVLRKVGEGYVFEKGLSRVYFAKDLFSEIVKNLESLSKQ ESTVHTITAKVDDENAPRDOEFYDAYNLLTEAHKALFENKGRNSDFQDLKDLRLND BSWTKELVADLAFAPITPHERLGNKNSQIETEDFVIAQADKYTTSDGYIFDE HD11SDEGDAVVTTHMGSHWIGKDSLSDEKVAAYTEKKGILPPSPDADVKANPT GDSAAAIYNRVYKGRKIPILVRLPYVEHTVEKGNLIIIPKHQVHNIKFADFDDHTY KAPNGYTLDELFIATIKYVVEPDERPHSNDGKNASHEVLGKDKHSDPKNFKADBE PVEETPAPEVPQVETEKVRAQLKEAEVLLAKVITDSSUKANATETLAGLRNLLTIQIM DNNSIMAEALALLKGSNPSVSKKIN" complement (8122. .9855) /gene="ptsI" /note="synonym: spr1062" complement (8122. .9855) /gene="ptsI" /EC_number="2.7.3.9" /codon_start=1 /transl_table=11 /label=spr1062 /product="PTS enzyme I" /protein_id="AAK99866.1" /db_xref="GI:15458684" /translation="MTEMLKGIAGSDGVAVAKAYLLVQPDLSFETITVEDTNAEEARL DAALQASODELSVIREKAVGTGEEAAQVFDHLVLPADPEMISOIKETIRAKKNAAE AGLKEVDMFITIPEGMEDNPMYQRAADIROVTKRVLANLLGKLPNPASINEEVIV IAHDLTPSDTAQIDKRFVYAFVNIIGGRTSHSAIMARTLEIAAIVLGTNNITETIVKDG ILAVNGITGEVLIINPTDEQAEFKAAAGEAYAKQAEWALLKDAQVTVDAGKHFLAAN IGTPKDEGVNNGAEAVGLYRTEFLYMDSPQPTDEDEQTEAYKAVLEGNMKEPVVR TMDIGDEKELPYFDPHENMPLGFRALRISISSETDAMFRTOIRALLRASVHGQLRI MFPVALLKEFKAAKAVDEEKNALIAEGVADNIIQVGMIEIIPAAAMLADQFAKEV DPFSGTNDLLIOYTWAAADRMEKQSVLYQVYNPISLRLLINNIKAHAEGKWAGMGE MAGDOQAVPLLVMGLDPEFSATSVLRTSLMKKLDITAKMEEYANRALTEGCTMBEV LELQKEYVNFDP" complement (9861. .10124) /gene="ptsH" /note="synonym: spr1063" complement (9861. .10124) /gene="ptsH" /codon_start=1 /transl_table=11 /label=spr1063 /product="Histidine-containing phosphocarrier protein of the PTS" /protein_id="AAK99867.1" /db_xref="GI:15458685" /translation="WASKDPHVVAETGCIHARPATLLVQTRASKFASDITLEYKGSVNL KSMGVMSLGVGGQADVTISAEAGDADDAIAAISSETMEKEGLA" | | | | | | | | | | | |
| ORIGIN | | | | | | | | | | | |
| Query Match | | 68.0% | Score | 1624; | DB | 1; | Length | 10320; | | | |
| Best Local Similarity | | 99.4% | Pred. | No. | 0; | Mismatches | 15; | Indels | 0; | Gaps | 0; |
| Matches | | 2374; | Conservative | 0; | | | | | | | |
| QY | 1 | TTCTTACAGTGTGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA | 60 | | | | | | | | |
| Db | 7808 | TTCTTACAGTGTGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA | 7749 | | | | | | | | |
| QY | 61 | TATAGATGGAACAAACGACGCGCAAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAA | 120 | | | | | | | | |
| Db | 7748 | TATAGATGGAACAAACGACGCGCAAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAA | 7689 | | | | | | | | |
| QY | 121 | CGGTGAGGAATCAATCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTAC | 180 | | | | | | | | |
| Db | 7688 | CGGTGAGGAATCAATCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTAC | 7629 | | | | | | | | |
| QY | 181 | TTACATGGGACCACTATCATTTATTAATGGAAGGTTCCTTTATGACGCTATCATCAG | 240 | | | | | | | | |
| Db | 7628 | TTACATGGGACCACTATCATTTATTAATGGAAGGTTCCTTTATGACGCTATCATCAG | 7569 | | | | | | | | |
| QY | 241 | TGAAGAATTAATCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATGTTAATGA | 300 | | | | | | | | |
| Db | 7568 | TGAAGAATTAATCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATGTTAATGA | 7509 | | | | | | | | |
| QY | 301 | GGTCAAGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC | 360 | | | | | | | | |

| | | | |
|----|------|---|------|
| Db | 7508 | GGTCAAGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC | 7449 |
| QY | 361 | TGCCACCGCATACAGTCGCTACAAAAGAGGAAATCAATGCACAAAAACAAGACATAG | 420 |
| Db | 7448 | TGCCACCGCATACAGTCGCTACAAAAGAGGAAATCAATGCACAAAAACAAGACATAG | 7389 |
| QY | 421 | TCAACATCGTGAAGGTGGAACTCCAAAGAAACGATGGTGTGTTCCTTGGGACGTTGCGA | 480 |
| Db | 7388 | TCAACATCGTGAAGGTGGAACTCCAAAGAAACGATGGTGTGTTCCTTGGGACGTTGCGA | 7329 |
| QY | 481 | AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC | 540 |
| Db | 7328 | AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC | 7269 |
| QY | 541 | TGCTGATGCTTATATCGTTCCTCATCGGATCATTTACCATTTACATTCCTAAGAATGAGTT | 600 |
| Db | 7268 | TGCTGATGCTTATATCGTTCCTCATCGGATCATTTACCATTTACATTCCTAAGAATGAGTT | 7209 |
| QY | 601 | ATCAGCTAGCGAGTTGGCTGTGCGAAGCCTTCCTATCTGTCGAGGAAATCTGTCAA | 660 |
| Db | 7208 | ATCAGCTAGCGAGTTGGCTGTGCGAAGCCTTCCTATCTGTCGAGGAAATCTGTCAA | 7149 |
| QY | 661 | TTCAAGAACCTATCGCCGACAAAATAGCGATACACTTCAAGAAACAACTGGGTACTCTTC | 720 |
| Db | 7148 | TTCAAGAACCTATCGCCGACAAAATAGCGATACACTTCAAGAAACAACTGGGTACTCTTC | 7089 |
| QY | 721 | TGTAAACCAATCCAGGAACCTACAAATACTAACCAAGCAACACAGCAACACTAACAGTCA | 780 |
| Db | 7088 | TGTAAACCAATCCAGGAACCTACAAATACTAACCAAGCAACACAGCAACACTAACAGTCA | 7029 |
| QY | 781 | AGCAAGTCAAGGTAATGACATTTGATAGTCTCTTTGAAAACAGCTCTACAAATCGCTTGG | 840 |
| Db | 7028 | AGCAAGTCAAGGTAATGACATTTGATAGTCTCTTTGAAAACAGCTCTACAAATCGCTTGG | 6969 |
| QY | 841 | TCAACGACATGATAGATCTGATGGCTGTGCTTTGATCCAGCAACAACTACAACTCGAAC | 900 |
| Db | 6968 | TCAACGACATGATAGATCTGATGGCTGTGCTTTGACCCGCGCAATCAAGTTCGAAC | 6909 |
| QY | 901 | AGCTAGAGGTGTGTCAGTGCACACGAGATCATTTACCACTTCATCCCTTACTCTCAAT | 960 |
| Db | 6908 | AGCTAGAGGTGTGTCAGTGCACACGAGATCATTTACCACTTCATCCCTTACTCTCAAT | 6849 |
| QY | 961 | GTCTGAATTTGGAAGAACGAATCGCTGATATTTCCCTTCCTGTTATCGTTCAACCACTG | 1020 |
| Db | 6848 | GTCTGAATTTGGAAGAACGAATCGCTGATATTTCCCTTCCTGTTATCGTTCAACCACTG | 6789 |
| QY | 1021 | GGTACCGAGATCAAGGCCGAAACCAACCACTCCACACCGACTCCGGAACCTAGTCCAGG | 1080 |
| Db | 6788 | GGTACCGAGATCAAGGCCGAAACCAACCACTCCACACCGACTCCGGAACCTAGTCCAGG | 6729 |
| QY | 1081 | CCGCAACCTGCAACCAATCTTAAATAGACTCAAATTTCTTTTGGTTAGTCACTGCT | 1140 |
| Db | 6728 | CCGCAACCTGCAACCAATCTTAAATAGACTCAAATTTCTTTTGGTTAGTCACTGCT | 6669 |
| QY | 1141 | ACGAAAAAGTTGGGGAAGGATATGTTATTCGAAGAAAGGCGATCTCTGTTATGCTTTTC | 1200 |
| Db | 6668 | ACGAAAAAGTTGGGGAAGGATATGTTATTCGAAGAAAGGCGATCTCTGTTATGCTTTTC | 6609 |
| QY | 1201 | GAAGAATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG | 1260 |
| Db | 6608 | GAAGAATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG | 6549 |
| QY | 1261 | TGTTTTCACACATTTAACTGCTAAAAAGAAAAATGTTGCTCCTCGTGACCAAGAAATTTA | 1320 |
| Db | 6548 | TGTTTTCACACATTTAACTGCTAAAAAGAAAAATGTTGCTCCTCGTGACCAAGAAATTTA | 6489 |
| QY | 1321 | TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTGTTTGNAAATAAGGGTCG | 1380 |
| Db | 6488 | TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTGTTTGNAAATAAGGGTCG | 6429 |
| QY | 1381 | TAAATCTGATTTCCAGCCTTTAGACAAATTTATAGAACGCTTGAATGATGATGACATAA | 1440 |

[illegible]

| | | | | | | | |
|----|--------|--|--------|----|--------|---|--------|
| Qy | 241 | TGAAGATTACTCATGAAGATCCAACTATAAGCTTAAAGATGAGGATATTGTTAAATGA | 300 | Qy | 1321 | TGATAAAGCATATATACTGTAACTAGAGCTCATAAAGCCTTGTGTTGNAATAAAGGTCG | 1380 |
| Db | 106710 | TGAAGATTACTCATGAAGATCCAACTATAAGCTTAAAGATGAGGATATTGTTAAATGA | 106651 | Db | 105630 | TGATAAAGCATATATACTGTAACTAGAGCTCATAAAGCCTTGTGTTGNAATAAAGGTCG | 105571 |
| Qy | 301 | GGTCAAGGGTGGATATGTTATCAAGTGTAGTGAATAATCTATGTTTACCTTTAAGGATGC | 360 | Qy | 1381 | TAAATCTGATTTCCAAAGCCTTAGACAAATATTAGAACGCTTGAATGATCGACTAA | 1440 |
| Db | 106650 | GGTCAAGGGTGGATATGTTATCAAGTGTAGTGAATAATCTATGTTTACCTTTAAGGATGC | 106591 | Db | 105570 | TAAATCTGATTTCCAAAGCCTTAGACAAATATTAGAACGCTTGAATGATCGACTAA | 105511 |
| Qy | 361 | TGCCACCGGATAACGTCCTGACAAAAGAGGAAATCAATGCAAAAAACAAGACATAG | 420 | Qy | 1441 | TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTCTAGCACCATAATACCCATCCAGAGCG | 1500 |
| Db | 106590 | TGCCACCGGATAACGTCCTGACAAAAGAGGAAATCAATGCAAAAAACAAGACATAG | 106531 | Db | 105510 | TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTCTAGCACCATAATACCCATCCAGAGCG | 105451 |
| Qy | 421 | TCAACATCGTGAAGTGGAACTCCCAAGAACGATGGTGTGCTTGGCACGTTGCGCA | 480 | Qy | 1501 | ACTTGGCAAAACCAAAATCTCAAATTTGAGTATCTCAAGACGAAAGTTCTGATTTCTCAAT | 1560 |
| Db | 106530 | TCAACATCGTGAAGTGGAACTCCCAAGAACGATGGTGTGCTTGGCACGTTGCGCA | 106471 | Db | 105450 | ACTTGGCAAAACCAAAATCTCAAATTTGAGTATCTCAAGACGAAAGTTCTGATTTCTCAAT | 105391 |
| Qy | 481 | AGGAGCTTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC | 540 | Qy | 1561 | AGCTGATTAAGTATACAACTGATGATGTTTACATTTTGTGATGAACATGATATATCAGTCA | 1620 |
| Db | 106470 | AGGAGCTTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC | 106411 | Db | 105390 | AGCTGATTAAGTATACAACTGATGATGTTTACATTTTGTGATGAACATGATATATCAGTCA | 105331 |
| Qy | 541 | TGGTGATGCTTATATCGTTCCTCATGGAGATCATTAACATTAATTCCTTAAGATGAGTT | 600 | Qy | 1621 | TGAAGAGATGATCATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAAAAGATAG | 1680 |
| Db | 106410 | TGGTGATGCTTATATCGTTCCTCATGGAGATCATTAACATTAATTCCTTAAGATGAGTT | 106351 | Db | 105330 | TGAAGAGATGATCATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAAAAGATAG | 105271 |
| Qy | 601 | ATCAGCTAGCGATTTGGCTGCTGCAGAACCTTCTCTATCTGCTGCGAGGAATCTGTCAA | 660 | Qy | 1681 | CTTTCTTGATAAGGAAAAAGTTGCAAGCTCAAGCCTTATCTAAAGAAAAAGGATCTCTACC | 1740 |
| Db | 106350 | ATCAGCTAGCGATTTGGCTGCTGCAGAACCTTCTCTATCTGCTGCGAGGAATCTGTCAA | 106291 | Db | 105270 | CTTTCTTGATAAGGAAAAAGTTGCAAGCTCAAGCCTTATCTAAAGAAAAAGGATCTCTACC | 105211 |
| Qy | 661 | TTCAAGAACCTATCGCGACAAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC | 720 | Qy | 1741 | TCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGAGATAGTGCAGCAGCTATTTTA | 1800 |
| Db | 106290 | TTCAAGAACCTATCGCGACAAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC | 106231 | Db | 105210 | TCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGAGATAGTGCAGCAGCTATTTTA | 105151 |
| Qy | 721 | TGTAAAGCAATCCAGAACTACAAATCTAACTAAGCAACAGCAACCACTAACTAGTCA | 780 | Qy | 1801 | CAATCGTGTGAAAGGGAAAAACGAATCCACTCGTTCGACTTCCATATATGTTGGAGCA | 1860 |
| Db | 106230 | TGTAAAGCAATCCAGAACTACAAATCTAACTAAGCAACAGCAACCACTAACTAGTCA | 106171 | Db | 105150 | CAATCGTGTGAAAGGGAAAAACGAATCCACTCGTTCGACTTCCATATATGTTGGAGCA | 105091 |
| Qy | 781 | AGCAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 840 | Qy | 1861 | TACAGTTGAGGTTTAAAGACGGTAAATTTGATTTTCTCTATAGGATCATTTACATATAT | 1920 |
| Db | 106170 | AGCAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 106111 | Db | 105090 | TACAGTTGAGGTTTAAAGACGGTAAATTTGATTTTCTCTATAGGATCATTTACCATATAT | 105031 |
| Qy | 841 | TCAACGACATGTAGAACTGTGATGGCCCTGTCTTTGATCCAGCAAAATCACAAGTCAAC | 900 | Qy | 1921 | TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 1980 |
| Db | 106110 | TCAACGACATGTAGAACTGTGATGGCCCTGTCTTTGATCCAGCAAAATCACAAGTCAAC | 106051 | Db | 105030 | TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 104971 |
| Qy | 901 | AGCTAGAGTGTGAGTGCACACAGGATGATGATGATGATGATGATGATGATGATGATGAT | 960 | Qy | 1981 | TTTGTGTCGACGATTAAGTACTACGTAGAACACCTCGACCAAGCTCCACATTTCTAATGA | 2040 |
| Db | 106050 | AGCTAGAGTGTGAGTGCACACAGGATGATGATGATGATGATGATGATGATGATGATGAT | 105991 | Db | 104970 | TTTGTGTCGACGATTAAGTACTACGTAGAACACCTCGACCAAGCTCCACATTTCTAATGA | 104911 |
| Qy | 961 | GTCTGAATTTGGAGAACGATCGCTGATATTTCCCTTCTGTTATCTCAACCATTTG | 1020 | Qy | 2041 | TGGATGGGCAATGCGCAGTGAGCATGTTTAGCAAGAAAGACCAAGTGAAGATCCAA | 2100 |
| Db | 105990 | GTCTGAATTTGGAGAACGATCGCTGATATTTCCCTTCTGTTATCTCAACCATTTG | 105931 | Db | 104910 | TGGATGGGCAATGCGCAGTGAGCATGTTTAGCAAGAAAGACCAAGTGAAGATCCAA | 104851 |
| Qy | 1021 | GGTACAGATTTCAAGGCCAGAACCAACAGTGCACAAACCGACTCCGGAACCTTAGTCCAG | 1080 | Qy | 2101 | TAAAGATTTCAAGCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG | 2160 |
| Db | 105930 | GGTACAGATTTCAAGGCCAGAACCAACAGTGCACAAACCGACTCCGGAACCTTAGTCCAG | 105871 | Db | 104850 | TAAAGATTTCAAGCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG | 104791 |
| Qy | 1081 | CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGGT | 1140 | Qy | 2161 | TCAAGTAGAGACTGAAAAAGTGAAGCCCAACTCAAGAGAGCAAGAGTGTGTTGCTGCGAA | 2220 |
| Db | 105870 | CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGGT | 105811 | Db | 104790 | TCAAGTAGAGACTGAAAAAGTGAAGCCCAACTCAAGAGAGCAAGAGTGTGTTGCTGCGAA | 104731 |
| Qy | 1141 | ACGAAAGTTGGGGAAGGATATGTTTCAAGAAAAAGGAGGATCTCTGTTATGCTTTTGC | 1200 | Qy | 2221 | AGTAAAGGATTTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGTTTGAAGATTA | 2280 |
| Db | 105810 | ACGAAAGTTGGGGAAGGATATGTTTCAAGAAAAAGGAGGATCTCTGTTATGCTTTTGC | 105751 | Db | 104730 | AGTAAAGGATTTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGTTTGAAGATTA | 104671 |
| Qy | 1201 | GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTATCAAAAAACAAGAG | 1260 | Qy | 2281 | TTTGACTCTTCAAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 2340 |
| Db | 105750 | GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTATCAAAAAACAAGAG | 105691 | Db | 104670 | TTTGACTCTTCAAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 104611 |
| Qy | 1261 | TGTTTCAACACTTAACTGCTAAAAAGAAATGTTGCTTCTCGTGACCAAGATTTTA | 1320 | Qy | 2341 | GTGTTTAAAAAGGAAGTAAATCTTCTCATCTGTGTAAGTAAAGAAAAATAAAC | 2389 |
| Db | 105690 | TGTTTCAACACTTAACTGCTAAAAAGAAATGTTGCTTCTCGTGACCAAGATTTTA | 105631 | Db | 104610 | GTGTTTAAAAAGGAAGTAAATCTTCTCATCTGTGTAAGTAAAGAAAAATAAAC | 104562 |

| | | | |
|----|------|--|------|
| QY | 961 | GTCCTGAATTTGAGAAACGAATCGCTGCTGATTATTTATCCCTTCTGGTTATCGTTCAAAACCATTG | 1020 |
| DB | 6777 | GTCTGAATTTGGAAGAAACGAATCGCTGCTGATTATTTATCCCTTCTGGTTATCGTTCAAAACCATTG | 6718 |
| QY | 1021 | GGTACCAGATTTCAAGGCCAGAACCAACCAAGTCCACAACCGACTCGGNAACCTAGTCCAGG | 1080 |
| DB | 6717 | GGTACCAGATTTCAAGGCCAGAACCAACCAAGTCCACAACCGACTCGGNAACCTAGTCCAGG | 6658 |
| QY | 1081 | CCCGCAACCTGCACCAAATCTTAAAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGGT | 1140 |
| DB | 6657 | CCCGCAACCTGCACCAAATCTTAAAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGGT | 6598 |
| QY | 1141 | ACGAAAAGTTGGGGAAGGATGATTTCGAAGAAAGGCAATCTCTCGTTATGCTTTTGC | 1200 |
| DB | 6597 | ACGAAAAGTTGGGGAAGGATGATTTCGAAGAAAGGCAATCTCTCGTTATGCTTTTGC | 6538 |
| QY | 1201 | GAAAGATTTACCATCTGAAACTGTTTAAAAATCTTTGAAAGCAAGTTATCMAAACHAAGAG | 1260 |
| DB | 6537 | GAAAGATTTACCATCTGAAACTGTTTAAAAATCTTTGAAAGCAAGTTATCMAAACHAAGAG | 6478 |
| QY | 1261 | TGTTTTCACACTTTAACTGCTAAAAAGAAAAATGTTGCTCCTCGTGACCAAGAATTTTA | 1320 |
| DB | 6477 | TGTTTTCACACTTTAACTGCTAAAAAGAAAAATGTTGCTCCTCGTGACCAAGAATTTTA | 6418 |
| QY | 1321 | TGATAAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCTTGTTGNAAATTAAGGTCG | 1380 |
| DB | 6417 | TGATAAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCTTGTTGNAAATTAAGGTCG | 6358 |
| QY | 1381 | TAATCTCTGATTTCCAAGCCTTTAGACAATATTATAGAACGCTTGGAATGATGAATCGACTAA | 1440 |
| DB | 6357 | TAATCTCTGATTTCCAAGCCTTTAGACAATATTATAGAACGCTTGGAATGATGAATCGACTAA | 6298 |
| QY | 1441 | TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTCCTAGCAACCAATACCOCATCCAGAGCG | 1500 |
| DB | 6297 | TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTCCTAGCAACCAATACCOCATCCAGAGCG | 6238 |
| QY | 1501 | ACTTGGCAAAACCAAAATCTCAAATGAGTATCTGAAGACGAAGTTCGTATTGCTCAAT | 1560 |
| DB | 6237 | ACTTGGCAAAACCAAAATCTCAAATGAGTATCTGAAGACGAAGTTCGTATTGCTCAAT | 6178 |
| QY | 1561 | AGCTGATAAGTATACAACGTCAGATGTTACATTTTGTAGAACATGATATAATCACTGA | 1620 |
| DB | 6177 | AGCTGATAAGTATACAACGTCAGATGTTACATTTTGTAGAACATGATATAATCACTGA | 6118 |
| QY | 1621 | TGAAGGAGATGTCATATGTTAAAGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG | 1680 |
| DB | 6117 | TGAAGGAGATGTCATATGTTAAAGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG | 6058 |
| QY | 1681 | CCTTTCTGATTAAGGAAAAAGTTCCAGCTCAAGCCTATACCTAAAGAAAAAGGTATCCTACC | 1740 |
| DB | 6057 | CCTTTCTGATTAAGGAAAAAGTTCCAGCTCAAGCCTATACCTAAAGAAAAAGGTATCCTACC | 5998 |
| QY | 1741 | TCCATCTCCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATGTCGACGAGCTATTTA | 1800 |
| DB | 5997 | TCCATCTCCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATGTCGACGAGCTATTTA | 5938 |
| QY | 1801 | CBAATCTGTTGAAAGGGGAAAAACGAATTCBACTCGCTTCCATATATGCTTGGAGCA | 1860 |
| DB | 5937 | CAATCTGTTGAAAGGGGAAAAACGAATTCBACTCGCTTCCATATATGCTTGGAGCA | 5878 |
| QY | 1861 | TACAGTTGAGGTTAAAAACGGTAAATTTGATATTCCTCATAAGGATCATTAACATAAT | 1920 |
| DB | 5877 | TACAGTTGAGGTTAAAAACGGTAAATTTGATATTCCTCATAAGGATCATTAACATAAT | 5818 |
| QY | 1921 | TAAATTTGCTTGGTTTGGATGATCACAATACAAGCTCCAAATGGCTATACCTTGGAGA | 1980 |
| DB | 5817 | TAAATTTGCTTGGTTTGGATGATCACAATACAAGCTCCAAATGGCTATACCTTGGAGA | 5758 |
| QY | 1981 | TTTGGTTGGGACGATTAAGTACTCGTAGAACACCTCGACGAACGTCACATTTCTTAATGA | 2040 |
| DB | 5757 | TTTGGTTGGGACGATTAAGTACTCGTAGAACACCTCGACGAACGTCACATTTCTTAATGA | 5698 |
| QY | 2041 | TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCCACAGTGAAGATCCAAA | 2100 |

| | | | | | | | | | | |
|---|--|--|-----------------|--------|--|--|--|--|--|--|
| Db | 5697 | TGGATGGGCAATGCCAGTGGCATGTGTAGGCAAGAAGACCA | CAGTGAAGATCCAAA | 5638 | | | | | | |
| QY | 2101 | TAAGAATTTCAAAGCGGATGAAGAGCGAGTAGAGGAAACACCTGCTGAGCGAGAGTCCC | 2160 | | | | | | | |
| Db | 5637 | TAAGAATTTCAAAGCGGATGAAGAGCGAGTAGAGGAAACACCTGCTGAGCGAGAGTCCC | 5578 | | | | | | | |
| QY | 2161 | TCAAGTAGAGACTGAAGAAAGTAGAGCCCACTCAAGAAAGCAGAAAGTTTGGTTGCGAA | 2220 | | | | | | | |
| Db | 5577 | TCAAGTAGAGACTGAAGAAAGTAGAGCCCACTCAAGAAAGCAGAAAGTTTGGTTGCGAA | 5518 | | | | | | | |
| QY | 2221 | AGTAACCGGATTTAGTCTGAAGCCCAATGCAACAGAAACTTAGCTGGTTTACGAAATAA | 2280 | | | | | | | |
| Db | 5517 | AGTAACCGGATTTAGTCTGAAGCCCAATGCAACAGAAACTTAGCTGGTTTACGAAATAA | 5458 | | | | | | | |
| QY | 2281 | TTTGACTCTTCAAATTTATGGATTAACATAGTATCATGGCAGAGCAGAAAAATTA | 2340 | | | | | | | |
| Db | 5457 | TTTGACTCTTCAAATTTATGGATTAACATAGTATCATGGCAGAGCAGAAAAATTA | 5398 | | | | | | | |
| QY | 2341 | GTTCTTAAAAGGAAGTAATCTTTCATCTCTAGTAGGAGGAAAAATAAC | 2389 | | | | | | | |
| Db | 5397 | GTTCTTAAAAGGAAGTAATCTTTCATCTCTAGTAGGAGGAAAAATAAC | 5349 | | | | | | | |
| RESULT 11 | | | | | | | | | | |
| AX571763/c | | | | | | | | | | |
| LOCUS | AX571763 | 349980 bp | DNA | linear | | | | | | |
| DEFINITION | Sequence 4982 from Patent WO02077021. | | | | | | | | | |
| ACCESSION | AX571763 | | | | | | | | | |
| VERSION | AX571763.1 | GI:26003955 | | | | | | | | |
| KEYWORDS | | | | | | | | | | |
| SOURCE | Streptococcus pneumoniae | | | | | | | | | |
| ORGANISM | Streptococcus pneumoniae | | | | | | | | | |
| REFERENCE | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; | | | | | | | | | |
| AUTHORS | Masignani, V., Tettelin, H. and Fraser, C. | | | | | | | | | |
| TITLE | Streptococcus pneumoniae proteins and nucleic acids | | | | | | | | | |
| JOURNAL | Patent: WO 02077021-A 4982 03-OCT-2002; | | | | | | | | | |
| FEATURES | Chiron Spa (IR); THE INSTITUTE FOR GENOMIC RESEARCH (US) | | | | | | | | | |
| ORIGIN | Location/Qualifiers | | | | | | | | | |
| | 1..349980 | | | | | | | | | |
| /organism="Streptococcus pneumoniae" | | | | | | | | | | |
| /mol_type="unassigned DNA" | | | | | | | | | | |
| /db_xref="taxon:1313" | | | | | | | | | | |
| /note="seq 4979 too long: 2.162.598 bases-replaced by following seq.-seq 4979: from 0.000.001 to 0.349.980-seq 4980: from 0.300.001 to 0.649.980-seq 4981: from 0.600.001 to 0.949.980-seq 4982: from 0.900.001 to 1.249.980-seq 4983: from 1.200.001 to 1.549.980-seq 4984: from 1.500.001 to 1.849.980-seq 4985: from 1.800.001 to 2.149.980-seq 4986: from 2.100.001 to 2.162.598" | | | | | | | | | | |
| Query Match 97.9%; Score 2338; DB 6; Length 349980; | | | | | | | | | | |
| Best Local Similarity 100.0%; Pred. No. 0; | | | | | | | | | | |
| Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | | | | | | | | | |
| QY | 1 | TTCTTACAGATTGGGACGTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA | 60 | | | | | | | |
| Db | 106950 | TTCTTACAGATTGGGACGTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA | 106891 | | | | | | | |
| QY | 61 | TATAGATCGAAAAAAGGCGACGCAAAAAACGGAGAAATTCGACTTCCTGATGAGTTAGCA | 120 | | | | | | | |
| Db | 106890 | TATAGATGAAAAAAGGCGACGCAAAAAACGGAGAAATTCGACTTCCTGATGAGTTAGCA | 120 | | | | | | | |
| QY | 121 | GGGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGACCAAGCCTATGTCAC | 180 | | | | | | | |
| Db | 106830 | GGGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGACCAAGCCTATGTCAC | 180 | | | | | | | |
| QY | 181 | TTACATCGGCACCACTATCAATTTAATGTTAAGGTTCCCTTATGAGCCTATCATCAG | 240 | | | | | | | |
| Db | 106770 | TTACATCGGCACCACTATCAATTTAATGTTAAGGTTCCCTTATGAGCCTATCATCAG | 106711 | | | | | | | |

| | | | |
|--|------------|---|------------------------|
| Db | 4853 | CAATCGTGTGAAGGGGAAACGAAATCCACTGTTGCACTTCATATATGTTGAGCA | 4912 |
| Qy | 1861 | TACAGTTGAGGTTAAACCGTAATTTGATTATTCCTCATAGGATCATACCAATAAT | 1920 |
| Db | 4913 | TACAGTTGAGGTTAAACCGTAATTTGATTATTCCTCATAGGATCATACCAATAAT | 4972 |
| Qy | 1921 | TAAATTTGCTTGGTTGATGATACACATACAGAGCTCCAAATGGCTATACCTTGGAGA | 1980 |
| Db | 4973 | TAAATTTGCTTGGTTGATGATACACATACAGAGCTCCAAATGGCTATACCTTGGAGA | 5032 |
| Qy | 1981 | TTTGTTCGCGAGTAATAGTACTAGTAGAACACCTGACGAACGTCACATTCCTAATGA | 2040 |
| Db | 5033 | TTTGTTCGCGAGTAATAGTACTAGTAGAACACCTGACGAACGTCACATTCCTAATGA | 5092 |
| Qy | 2041 | TGGATGGGGAATCGCAGTAGCATGTGTAGGCAAGAACACACAGTGAAGATCCAAA | 2100 |
| Db | 5093 | TGGATGGGGAATCGCAGTAGCATGTGTAGGCAAGAACACACAGTGAAGATCCAAA | 5152 |
| Qy | 2101 | TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTCTGAGCCAGAGTCCC | 2160 |
| Db | 5153 | TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTCTGAGCCAGAGTCCC | 5212 |
| Qy | 2161 | TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAGAAAGCAGAAGTTTGTCTTGGAA | 2220 |
| Db | 5213 | TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAGAAAGCAGAAGTTTGTCTTGGAA | 5272 |
| Qy | 2221 | AGTAACCGATCTAGTCTGAAGCGAATGCAACGAACTCTAGCTGGTTACGAATAA | 2280 |
| Db | 5273 | AGTAACCGATCTAGTCTGAAGCGAATGCAACGAACTCTAGCTGGTTACGAATAA | 5332 |
| Qy | 2281 | TTTGACTCTTCAAAATTTAGTATAACAATAGTATCATGCGAAGCAGAAATTTACTTGC | 2340 |
| Db | 5333 | TTTGACTCTTCAAAATTTAGTATAACAATAGTATCATGCGAAGCAGAAATTTACTTGC | 5392 |
| Qy | 2341 | GTGTGTTAAAGGAAGTAATCTTCTCATCTGTAAGTAAGCAAAAATAAAC | 2389 |
| Db | 5393 | GTGTGTTAAAGGAAGTAATCTTCTCATCTGTAAGTAAGCAAAAATAAAC | 5441 |
| RESULT 10 | | | |
| AE007418/c | | 10256 bp | DNA |
| LOCUS | | | |
| DEFINITION | AE007418 | Streptococcus pneumoniae TIGR4 | linear BCT 31-AUG-2001 |
| ACCESSION | AE005672 | | |
| VERSION | AE007418.1 | GI:14972649 | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| 1 (bases 1 to 10256) | | | |
| Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., | | | |
| Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., | | | |
| Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., | | | |
| Unayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., | | | |
| Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., | | | |
| McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., | | | |
| Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., | | | |
| Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and | | | |
| Fraser, C.M. | | | |
| Complete genome sequence of a virulent isolate of Streptococcus | | | |
| pneumoniae | | | |
| Science 293 (5529), 498-506 (2001) | | | |
| JOURNAL | | | |
| MEDLINE | | | |
| PUBMED | | | |
| 2 (bases 1 to 10256) | | | |
| Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., | | | |
| Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., | | | |
| Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., | | | |
| Unayam, L.A., White, O., Salzberg, S.L., Radune, D., Holtzapple, E., | | | |
| Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., | | | |

| | | | |
|---|--|--|--|
| TITLE | | | |
| JOURNAL | | | |
| FEATURES | | | |
| source | | | |
| 1. 10256 | | | |
| /organism="Streptococcus pneumoniae TIGR4" | | | |
| /mol_type="genomic DNA" | | | |
| /strain="TIGR4" | | | |
| /db_xref="taxon:170187" | | | |
| complement(100..702) | | | |
| /gene="SP1170" | | | |
| complement(100..702) | | | |
| /gene="SP1170" | | | |
| /note="identified by Glimmer2; putative" | | | |
| /codon_start=1 | | | |
| /transl_table=11 | | | |
| /product="hypothetical protein" | | | |
| /protein_id="AAK75279.1" | | | |
| /db_xref="GI:14972650" | | | |
| /translation="MSEVDFNEAVNYEFTSDTCLANSIYQSLEKFFDKNFSGDLIF | | | |
| TKPSLVKEGDYIGRRDQVDNLRVIGNIFPNLTNRKYSLNMRNGCMGDFPHDF | | | |
| DIYLDHVAKYAVQKNNIKEYVPLKRAILLHOENALYERFESNFDLLEKNYLTQW | | | |
| VSETFSEMDFNMPKNISEKIIIFERGSKMLNDLKSNNK" | | | |
| complement(760..1473) | | | |
| /gene="SP1171" | | | |
| complement(760..1473) | | | |
| /gene="SP1171" | | | |
| /note="identified by match to PFAM protein family HMM | | | |
| PF00702" | | | |
| /codon_start=1 | | | |
| /transl_table=11 | | | |
| /product="hydrolase, haloacid dehalogenase-like family" | | | |
| /protein_id="AAK75280.1" | | | |
| /db_xref="GI:14972651" | | | |
| /translation="MFYKLELLFDLDDHLLDFAAEDVALTQLLKESGVADIQAYKYY | | | |
| VPNKALWDELEKJISKQELVNTRFSLRFAHQKDGSLAQRFQYLAQKQQLS | | | |
| GAHLLDLSLIERDNLNATNGITATQTLAQSLAPYQVFISEQLTQKQDALF | | | |
| YXIQQIAGFSKERTIMIGDSLTADIQGNNAIDITWYNPHLENHTQAQPTVEY | | | |
| SYQQLDCLDKNLEKITP" | | | |
| complement(1724..1843) | | | |
| /gene="SP1172" | | | |
| complement(1724..1843) | | | |
| /gene="SP1172" | | | |
| /note="identified by Glimmer2; putative" | | | |
| /codon_start=1 | | | |
| /transl_table=11 | | | |
| /product="hypothetical protein" | | | |
| /protein_id="AAK75281.1" | | | |
| /db_xref="GI:14972652" | | | |
| /translation="MLYAVFPFNRSETIVFLNCESIKTDCGAILALETFKN" | | | |
| complement(1933..2307) | | | |
| /gene="SP1173" | | | |
| complement(1933..2307) | | | |
| /gene="SP1173" | | | |
| /note="identified by Glimmer2; putative" | | | |
| /codon_start=1 | | | |
| /transl_table=11 | | | |
| /product="conserved hypothetical protein" | | | |
| /protein_id="AAK75282.1" | | | |
| /db_xref="GI:14972653" | | | |
| /translation="MESIGLVISHSKHIAEGVWELISKVAKDVPTTYRTEGGGIG | | | |
| TSFEQDVRYSENPAITLAFDLGSAKNLKNVTFDSKSIIRNRPIVEGAYNAA | | | |
| LIQAGAEISVIQTQLAELEINK" | | | |
| complement(2729..5188) | | | |
| /gene="SP1174" | | | |
| complement(2729..5188) | | | |
| /gene="SP1174" | | | |
| /note="similar to GP:4249624; identified by sequence | | | |
| similarity; putative" | | | |
| /codon_start=1 | | | |

Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E., Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M. Direct Submission
Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

| | |
|--------|--|
| PF | 30-OCT-1997 JP 1998520718 |
| PR | 31-OCT-1996 US 60/029960 |
| PI | CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI STEVEN C BARASH. |
| PT | MICHAEL FANNON, BRIAN A DOUGHERTY |
| PC | C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19, |
| PC | C12N1/21, |
| PC | C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC G06F15/40 |
| CC | Strandedness: Double; |
| CC | Topology: Linear; |
| PH | Key Location/Qualifiers |
| FT | source 1..8195 /organism='Unidentified'. |
| FT | Location/Qualifiers |
| | 1..8195 /organism='unidentified' |
| | /mol_type='genomic DNA' |
| | /db_xref='taxon:32644' |
| ORIGIN | |
| | Query Match 97.9%; Score 2338; DB 6; Length 8195; |
| | Best Local Similarity 100.0%; Pred. No. 0; |
| | Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0; |
| QY | 1 TTCTTACGAGTTGGACTGTATCAAGCTAAGCGTTAAGGAAAAATAATCGTGTTTCCTA 60 |
| Db | 3053 TTCTTACGAGTTGGACTGTATCAAGCTAAGCGTTAAGGAAAAATAATCGTGTTTCCTA 3112 |
| QY | 61 TATAGATGGAAAAACAAGCGACGCAAAAACGGAGAAATTTCACCTCCTGATGAGGTTAGCAA 120 |
| Db | 3113 TATAGATGGAAAAACAAGCGACGCAAAAACGGAGAAATTTCACCTCCTGATGAGGTTAGCAA 3172 |
| QY | 121 GCCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACAAGGCTATGTCAC 180 |
| Db | 3173 GCCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACAAGGCTATGTCAC 3232 |
| QY | 181 TTCACATGGCGACGACATCATTAATTACAATGTTAGGTTTCCTTATGACGCTATCATCAG 240 |
| Db | 3233 TTCACATGGCGCACCATCATTAATTACAATGTTAGGTTTCCTTATGACGCTATCATCAG 3292 |
| QY | 241 TGAAGAATTACTCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATTCGTTAATGA 300 |
| Db | 3293 TGAAGAATTACTCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATTCGTTAATGA 3352 |
| QY | 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTTCCTTAAGGATGC 360 |
| Db | 3353 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAGGATGC 3412 |
| QY | 361 TGCCCAACGGATAAAGTCGTCGATCAAAAGAGGAATCAATCGACAAAAACAAGAGCATAG 420 |
| Db | 3413 TGCCCAACGGATAAAGTCGTCGATCAAAAGAGGAATCAATCGACAAAAACAAGAGCATAG 3472 |
| QY | 421 TCACATCGTGAAGGTGGAACACTCAAGAACGATGGTCTGTTGCCCTTGGCAGGTTCCGCA 480 |
| Db | 3473 TCACATCGTGAAGGTGGAACACTCAAGAACGATGGTCTGTTGCCCTTGGCAGGTTCCGCA 3532 |
| QY | 481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540 |
| Db | 3533 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 3592 |
| QY | 541 TGGTGATGCTTATATCGTTCCTCATGAGATCATATACATTACATTCCTTAAGATGAGTT 600 |
| Db | 3593 TGGTGATGCTTATATCGTTCCTCATGAGATCATATACATTACATTCCTTAAGATGAGTT 3652 |
| QY | 601 ATCAGCTAGCGAGTTGGCTGCTGAGAAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAA 660 |
| Db | 3653 ATCAGCTAGCGAGTTGGCTGCTGAGAAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAA 3712 |
| QY | 661 TTCAGAAACCTATCGCCGACAAATACGGATACACCTTCAAGAACAACTGGGTACCTTC 720 |
| Db | 3713 TTCAGAAACCTATCGCCGACAAATACGGATACACCTTCAAGAACAACTGGGTACCTTC 3772 |

| | | | |
|------|-----------|---|------|
| 3533 | AGAGCGCTA | CTACTACAGATGATGGTTATATCTTTAATGCTTCTGTATCATAGAGATAC | 3599 |
| Qy | 541 | TGTTGATGCTTATATCTGCTTCATCGAGATCAATTCATTACATTCCTAAGAATGAGTT | 600 |
| | 3593 | TGGTGATGCTTATATCGTTCTCATCGAGATCATTACCATTACATTCCTAAGAATGAGTT | 3652 |
| | 601 | ATCAGCTAGCGAGTTGGCTGTCGAGAAGCGTTTCCCTATCTGGTCGAGGAATCTGTCAAA | 660 |
| Db | 3653 | ATCAGCTAGCGAGTTGGCTGTCGAGAAGCGTTTCCCTATCTGGTCGAGGAATCTGTCAAA | 3712 |
| | 661 | TTCAAGAACCTATTCGCCGCAAAATACGGATAACACTTTCAAGAACAACTGGGGTACCTTC | 720 |
| | 3713 | TTCAAGAACCTATTCGCCGCAAAATACGGATAACACTTTCAAGAACAACTGGGGTACCTTC | 3772 |
| Qy | 721 | TGTAAGCAATCCAGGAATCAAAATCTAACAACAAGCAACAACAGCAAACTAACAGTCA | 780 |
| Db | 3773 | TGTAAGCAATCCAGGAATCAAAATCTAACAACAAGCAACAACAGCAAACTAACAGTCA | 3832 |
| | 781 | AGCAAGTCAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTCAAACTGGCTTTGAG | 840 |
| | 3833 | AGCAAGTCAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTCTACAACTGGCTTTGAG | 3892 |
| Qy | 841 | TCAACGACATGTAGAACTGATGCGCTTGCTTTGATCCAGCAAAATCAACAAGTCGAAC | 900 |
| Db | 3893 | TCAACGACATGTAGAACTGATGCGCTTGCTTTGATCCAGCAAAATCAACAAGTCGAAC | 3952 |
| | 901 | AGCTAGAGGTGTTGTCAGTGGCCACACGAGAGATCATTAACCTTCATCCCTTACTCTCAAA | 960 |
| | 3953 | AGCTAGAGGTGTTGTCAGTGGCCACACGAGAGATCATTAACCTTCATCCCTTACTCTCAAA | 4012 |
| Qy | 961 | GTCTGAATTTGGAAGAACGAATCGCTCGTATTAATTCCTCCCTTCGGTTATCGTTCAAA | 1020 |
| Db | 4013 | GTCTGAATTTGGAAGAACGAATCGCTCGTATTAATTCCTCCCTTCGGTTATCGTTCAAA | 4072 |
| | 1021 | GGTACCAGATTCAGGCCAGAACCAACAGTCCACACACCGACCTCCGAACTAGTCCAGG | 1080 |
| | 4073 | GGTACCAGATTCAGGCCAGAACCAACAGTCCACACACCGACCTCCGAACTAGTCCAGG | 4132 |
| Qy | 1081 | CCGCAACCTGCACAAATCTTAAATATAGACTCAAAATCTTCTTTGGTTAGTCAGTGGT | 1140 |
| Db | 4133 | CCGCAACCTGCACCAATCTTAAATATAGACTCAAAATCTTCTTTGGTTAGTCAGTGGT | 4192 |
| | 1141 | ACGAAAGTTGGGGAAGGATATGTAATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTGTC | 1200 |
| | 4193 | ACGAAAGTTGGGGAAGGATATGTAATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTGTC | 4252 |
| Qy | 1201 | GAAGATTTACCATCTCAACTGTTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG | 1260 |
| Db | 4253 | GAAGATTTACCATCTCAACTGTTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG | 4312 |
| | 1261 | TGTTTTCACACATTTAACTGCTAAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA | 1320 |
| | 4313 | TGTTTTCACACATTTAACTGCTAAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA | 4372 |
| Qy | 1321 | TGATAAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATTAAGGGTCG | 1380 |
| Db | 4373 | TGATAAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATTAAGGGTCG | 4432 |
| | 1381 | TAATTTCTGATTTCCAGCCCTTAGACAAATATTTAGACGCTTGTAATGATCAATCGACTAA | 1440 |
| | 4433 | TAATTTCTGATTTCCAGCCCTTAGACAAATATTTAGACGCTTGTAATGATCAATCGACTAA | 4492 |
| Qy | 1441 | TAAAGAAAAATTTGGTATGATGATTTATTTGGCATTCCTAGACCAATTAACCATCCAGAGCG | 1500 |
| Db | 4493 | TAAAGAAAAATTTGGTATGATGATTTATTTGGCATTCCTAGACCAATTAACCATCCAGAGCG | 4552 |
| | 1501 | ACTTTGGCAAAACCAAAATCTCAAAATTTAGTATACCTGAAGACGAAGTTTCGTATTCGTTCAAT | 1560 |
| | 4553 | ACTTTGGCAAAACCAAAATCTCAAAATTTAGTATACCTGAAGACGAAGTTTCGTATTCGTTCAAT | 4612 |
| Qy | 1561 | AGCTGATAAGTATACACAGTCAGATGGTTACATTTTTTGTAGAACATGATATAATCACTGTA | 1620 |
| Db | 4613 | AGCTGATAAGTATACACAGTCAGATGGTTACATTTTTTGTAGAACATGATATAATCACTGTA | 4672 |

| | | | |
|----|------|--|------|
| Qy | 1621 | TGAAGGAGATGCATATGTTAAAGCCTCATATGGGCCATATGTCATCTGGATTGGAAAAAGATAG | 1680 |
| Db | 4673 | TGAAGGAGATGCATATGTTAAAGCCTCATATGGGCCATATGTCATCTGGATTGGAAAAAGATAG | 4732 |
| Qy | 1681 | CCTTTCTGTAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGNAAGGATCCTTACC | 1740 |
| Db | 4733 | CCTTTCTGTAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGNAAGGATCCTTACC | 4792 |
| Qy | 1741 | TCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA | 1800 |
| Db | 4793 | TCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA | 4852 |
| Qy | 1801 | CAATCTGTTGAAAGGGGAAAAACAATTCACCTCGTTTCGACTTCCATATATATGTTGAGCA | 1860 |
| Db | 4853 | CAATCTGTTGAAAGGGGAAAAACAATTCACCTCGTTTCGACTTCCATATATATGTTGAGCA | 4912 |
| Qy | 1861 | TACAGTTGAGGTTTAAAAACGGTAATTTTGATTATTCCTCATAAAGGATCATTTACCATAATAT | 1920 |
| Db | 4913 | TACAGTTGAGGTTTAAAAACGGTAATTTTGATTATTCCTCATAAAGGATCATTTACCATAATAT | 4972 |
| Qy | 1921 | TAAATTGCTTGGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGGAAGA | 1980 |
| Db | 4973 | TAAATTGCTTGGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGGAAGA | 5032 |
| Qy | 1981 | TTTGTGTCGACGATTAAGTACTACGTAGAACACCCCTGACGAACGCTCCACATCTTAATGA | 2040 |
| Db | 5033 | TTTGTGTCGACGATTAAGTACTACGTAGAACACCCCTGACGAACGCTCCACATCTTAATGA | 5092 |
| Qy | 2041 | TGGATGGGGCAATGCCAGTGACATGTTTAGGCAGAAAGACCCACAGTGAAGATCCAAA | 2100 |
| Db | 5093 | TGGATGGGGCAATGCCAGTGACATGTTTAGGCAGAAAGACCCACAGTGAAGATCCAAA | 5152 |
| Qy | 2101 | TAAGAACCTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAAGTCCC | 2160 |
| Db | 5153 | TAAGAACCTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAAGTCCC | 5212 |
| Qy | 2161 | TCAAGTAGAGACTCAAAAAGTAGAAGCCCAACTCAAAAGAGCAGAAAGTTTTGCTTGCAGAA | 2220 |
| Db | 5213 | TCAAGTAGAGACTCAAAAAGTAGAAGCCCAACTCAAAAGAGCAGAAAGTTTTGCTTGCAGAA | 5272 |
| Qy | 2221 | AGTAAACGATTTCTAGTCTGAAGGCCATGCAACAGAAACTCTAGCTGGTTTACGAATAAA | 2280 |
| Db | 5273 | AGTAAACGATTTCTAGTCTGAAGGCCATGCAACAGAAACTCTAGCTGGTTTACGAATAAA | 5332 |
| Qy | 2281 | TTTGACTCTTCAAATATGATTAACATAGTATCATGCGCAGAAAGCAAAAATTTACTTGC | 2340 |
| Db | 5333 | TTTGACTCTTCAAATATGATTAACATAGTATCATGCGCAGAAAGCAAAAATTTACTTGC | 5392 |
| Qy | 2341 | GTTGTTAAAAAGGAGTAATCCTTTCATCTGTAAGTAAGGAAAAATAAAC | 2389 |
| Db | 5393 | GTTGTTAAAAAGGAGTAATCCTTTCATCTGTAAGTAAGGAAAAATAAAC | 5441 |

RESULT 9

BD003774

LOCUS

DEFINITION
ACCESSION

VERSION

KEYWORDS

SOURCE: ORGANIZATION

TABLE 1

REFERENCES

AUTHOR.

TITLE

JOURNAL

2000

COMMENT

BD003774 8195 bp DNA linear PAT 31-JAN-2002
Polynucleotide of *Streptococcus pneumoniae* and sequence.

BD003774
BD003774.1 GI:18631735

JP 2001501833-A/94.
unidentified

unclassified.
unclassified.
unclassified.

I (bases 1 to 8195)
Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C.,
Fannon M. and Deuchter, B. A

Polynucleotide of *Streptococcus pneumoniae* and sequence
Patent: JP 2001501833-A 94 13-FEB-2001.

Account of Experiments in Optics (1801)
HUMAN GENOME SCIENCES INC
QS Unidentified

PN JP 2001501833-A/94
PD 13-FEB-2001

Db 1098 CCCGCAACCTGCACCAAACTTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCTGGT 1157
Qy 1141 ACGAAAAGTTGGGAGGATATGTATTTCGAAGAAAGGCGCATCTCGTTATGTCTTTGC 1200
Db 1158 ACGAAAAGTTGGGAGGATATGTATTTCGAAGAAAGGCGCATCTCGTTATGTCTTTGC 1217
Qy 1201 GAAAGATTTTACCATCTGAAACCTGTTAAAAATCTTGAAGCAAGTATCAAAAACAAGAG 1260
Db 1218 GAAAGATTTTACCATCTGAAACCTGTTAAAAATCTTGAAGCAAGTATCAAAAACAAGAG 1277
Qy 1261 TGTTCACACATTTAACTGCTAAAAAGAAAATGTTGCTTCGTCGACCAAGAAATTTA 1320
Db 1278 TGTTCACACATTTAACTGCTAAAAAGAAAATGTTGCTTCGTCGACCAAGAAATTTA 1337
Qy 1321 TGATAAAGCATATAATCTGTAACTGAGGCTCAATAAGGCTTGTGTTGNAATAAGGCTG 1380
Db 1338 TGATAAAGCATATAATCTGTAACTGAGGCTCAATAAGGCTTGTGTTGNAATAAGGCTG 1397
Qy 1381 TAATCTGATTTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1440
Db 1398 TAATCTGATTTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1457
Qy 1441 TAAAGAAAATGCTAGATGATTTATTGGCATCTCTAGCACCAATTAACCATCCAGAGCG 1500
Db 1458 TAAAGAAAATGCTAGATGATTTATTGGCATCTCTAGCACCAATTAACCATCCAGAGCG 1517
Qy 1501 ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAGTTCGTTATGCTCAAT 1560
Db 1518 ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAGTTCGTTATGCTCAAT 1577
Qy 1561 AGCTGATAAGTATACAACTGATGATGTTTACATTTTGTGATGAACATGATATAATCAGTGA 1620
Db 1578 AGCTGATAAGTATACAACTGATGATGTTTACATTTTGTGATGAACATGATATAATCAGTGA 1637
Qy 1621 TGAAGGAGATGATATGTAACGCTCATATGAGGCTATGCTGCTGATTTGAAAAGATAG 1680
Db 1638 TGAAGGAGATGATATGTAACGCTCATATGAGGCTATGCTGCTGATTTGAAAAGATAG 1697
Qy 1681 CTTTCTGATAAGGAAAAGTTGACGCTCAAGCTATATCTAAAGAAAAGGTTATCTACC 1740
Db 1698 CTTTCTGATAAGGAAAAGTTGACGCTCAAGCTATATCTAAAGAAAAGGTTATCTACC 1757
Qy 1741 TCCATCTCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
Db 1758 TCCATCTCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1817
Qy 1801 CAATCGTGTCAAGGGGAAAACGAATTCACCTCGTTCCACTTCCATATATGTTGAGCA 1860
Db 1818 CAATCGTGTCAAGGGGAAAACGAATTCACCTCGTTCCACTTCCATATATGTTGAGCA 1877
Qy 1861 TACAGTTGAGGTTTAAAGCGGTAATTTGATTTTCTCTATAAGGATCATTAACATAAT 1920
Db 1878 TACAGTTGAGGTTTAAAGCGGTAATTTGATTTTCTCTATAAGGATCATTAACATAAT 1937
Qy 1921 TAAATTTGCTGTTTGTGATCACACATPACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
Db 1938 TAAATTTGCTGTTTGTGATCACACATPACAAAGCTCCAAATGGCTATACCTTGGAGA 1997
Qy 1981 TTTGTTTGGAGATTAAGTACTAGTAGAACACCCCTGACGACGCTCCACATTTCTAATGA 2040
Db 1998 TTTGTTTGGAGATTAAGTACTAGTAGAACACCCCTGACGACGCTCCACATTTCTAATGA 2057
Qy 2041 TGGATGGGGCAATGCCAGTGAGCTGTTTGTAGCAAGAAAGACCAAGTGAAGATCCAAA 2100
Db 2058 TGGATGGGGCAATGCCAGTGAGCTGTTTGTAGCAAGAAAGACCAAGTGAAGATCCAAA 2117
Qy 2101 TAAGAACTTTCAAAGCGGATGAAGACCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
Db 2118 TAAGAACTTTCAAAGCGGATGAAGACCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2177
Qy 2161 TCAGGTAGAGCTGAAGAAAGTAGAGCCCAACTCAAGAGCAGAGTTTTCGTCGAA 2220
Db 2178 TCAAGTAGAGCTGAAGAAAGTAGAGCCCAACTCAAGAGCAGAGTTTTCGTCGAA 2237

Qy 2221 AGTAACGGATTTCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAATAA 2280
Db 2238 AGTAACGGATTTCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAATAA 2297
Qy 2281 TTTGACTCTTCAAAATTTATGGATAACAATAGTATATCGGAGAGCAGAAAAATTTACTTGC 2340
Db 2298 TTTGACTCTTCAAAATTTATGGATAACAATAGTATATCGGAGAGCAGAAAAATTTACTTGC 2357
Qy 2341 GTTGTAAAAGGAGTAATCTTTCATCTGTAGTAAGGAAAAATAAAC 2389
Db 2358 GTTGTAAAAGGAGTAATCTTTCATCTGTAGTAAGGAAAAATAAAC 2406

RESULT 8
AR218862
LOCUS AR218862
DEFINITION Sequence 94 from patent
ACCESSION AR218862
VERSION AR218862.1
KEYWORDS GI:23319796
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 8195)
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,
Fannon, M.R., and Dougherty, B.A.
TITLE Streptococcus pneumoniae polynucleotides and sequences
JOURNAL Patent: US 6420135-A 94 16-JUL-2002;
FEATURES Location/Qualifiers
 1..8195
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
Query Match 97.9%; Score 2338; DB 6; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TTTCTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATTAATCGTTTCTTA 60
Db 3053 TTTCTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATTAATCGTTTCTTA 3112
Qy 61 TATAGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCCTGATGAGTTAGCAA 120
Db 3113 TATAGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCCTGATGAGTTAGCAA 3172
Qy 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db 3173 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 3232
Qy 181 TTCAATGCGGACCACTATCATTTACATGTTAAGGTTTCTTATGAGCTATCATCAG 240
Db 3233 TTCAATGCGGACCACTATCATTTACATGTTAAGGTTTCTTATGAGCTATCATCAG 3292
Qy 241 TGAAGAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTTCTTAATGA 300
Db 3293 TGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTTCTTAATGA 3352
Qy 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAAAATACTATGTTTACTTAAGGATGC 360
Db 3353 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAAAATACTATGTTTACTTAAGGATGC 3412
Qy 361 TGCCCAACCGGATAACGTCCTGACAAAAGAGGAATCAATCCGCAAAAAACAGAGCATAG 420
Db 3413 TGCCCAACCGGATAACGTCCTGACAAAAGAGGAATCAATCCGCAAAAAACAGAGCATAG 3472
Qy 421 TCAACATCGTGAAGGTGGAACCTCCAAAGAAACCATGGTGTCTGTTGCCCTGCGACGTTTCGCA 480
Db 3473 TCAACATCGTGAAGGTGGAACCTCCAAAGAAACCATGGTGTCTGTTGCCCTGCGACGTTTCGCA 3532
Qy 481 AGGACGCTATACCTACAGATGATGGTTATCTTTAATGCTTCTGATPATCATAGAGATAC 540

```
QY 1741 TCCATCTCCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTGCACAGCTATTTA 1800
Db 1890 TCCATCTCCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTGCACAGCTATTTA 1949
QY 1801 CAATCGTGTCAAAAGGGGAAAAAGCAATTCACATCGTTCGACTTCCTATATATGTTGAGCA 1860
Db 1950 CAATCGTGTGAAGGGGAAAAAGCAATTCACATCGTTCGACTTCCTATATATGTTGAGCA 2009
QY 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATATATTCCTCATPAAGATCAATTAACATATAT 1920
Db 2010 TACAGTTGAGGTTAAAAACGGTAATTTGATATATTCCTCATPAAGATCAATTAACATATAT 2069
QY 1921 TAAATTTGCTTGGTTTGCATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
Db 2070 TAAATTTGCTTGGTTTGCATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 2129
QY 1981 TTTGTTTGGACGATTAAGTACTACGTAGAACACCCCTGACGAACGCTCCACATTTCTAATGA 2040
Db 2130 TTTGTTTGGACGATTAAGTACTACGTAGAACACCCCTGACGAACGCTCCACATTTCTAATGA 2189
QY 2041 TGGATGGGGCAATGCCAGTGAGCATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100
Db 2190 TGGATGGGGCAATGCCAGTGAGCATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2249
QY 2101 TAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
Db 2250 TAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2309
QY 2161 TCAAGTAGAGACTGAAAAGGTAGAAGCCCAACTCAAAAGAGCAGAAAGTTTGTCTTGGCAA 2220
Db 2310 TCAAGTAGAGACTGAAAAGGTAGAAGCCCAACTCAAAAGAGCAGAAAGTTTGTCTTGGCAA 2369
QY 2221 AGTAACGGATCTAGCTGAAAGCCATGCAAGAAACTCTAGCTGGTTTACGAATATA 2280
Db 2370 AGTAACGGATCTAGCTGAAAGCCATGCAAGAAACTCTAGCTGGTTTACGAATATA 2429
QY 2281 TTTGACTCTTCAATTTATGATACAAATAGTATCATGCGAAGCAGAAATAATTTACTTGC 2340
Db 2430 TTTGACTCTTCAATTTATGATACAAATAGTATCATGCGAAGCAGAAATAATTTACTTGC 2489
QY 2341 GTTGTTAAAGGAAGTAACTCTTCATCTGTAAGTAAAGGAAAAATAAAC 2389
Db 2490 GTTGTTAAAGGAAGTAACTCTTCATCTGTAAGTAAAGGAAAAATAAAC 2538

RESULT 7
AX569139 2406 bp DNA linear PAT 29-NOV-2002
LOCUS Sequence 2347 from Patent WO02077021.
DEFINITION AX569139
ACCESSION AX569139
VERSION AX569139.1 GI:26002636
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.

REFERENCE
1 Masignani, V., Tettelin, H. and Fraser, C.
AUTHORS Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 2347 03-OCT-2002;
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
Location/Qualifiers
source 1..2406
/organism="Streptococcus pneumoniae"
/mol_type="unassigned DNA"
/db_xref="taxon:1313"

ORIGIN
Query Match 97.9%; Score 2338; DB 6; Length 2406;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 60
```

```
Db 18 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 77
QY 61 TATAGATGGAACAAACAGCGACGCAAAAAACGGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
Db 78 TATAGATGGAACAAACAGCGACGCAAAAAACGGAGAAATTTGACTCCTGATGAGGTTAGCAA 137
QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 180
Db 138 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 197
QY 181 TTCAATATGCGGACCACTATCATTTATCAATGTAAGGTTCTTTATGACGCTATCATCAG 240
Db 198 TTCAATATGCGGACCACTATCATTTATCAATGTAAGGTTCTTTATGACGCTATCATCAG 257
QY 241 TGAAGAAATTTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAATGA 300
Db 258 TGAAGAAATTTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAATGA 317
QY 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGGATGC 360
Db 318 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGGATGC 377
QY 361 TGCCCAACGCGGATAACGTCCTGTAACAAAGAGGAAATCAATTCGAACAAAAACAAGAGCATAG 420
Db 378 TGCCCAACGCGGATAACGTCCTGTAACAAAGAGGAAATCAATTCGAACAAAAACAAGAGCATAG 437
QY 421 TCAACATCTGTAAGGTGGAACTCAAGAAACGATGGTCTGTTGCCCTTGGCACTTGGCA 480
Db 438 TCAACATCTGTAAGGTGGAACTCAAGAAACGATGGTCTGTTGCCCTTGGCACTTGGCA 497
QY 481 AGGACGCTATCTACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db 498 AGGACGCTATCTACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 557
QY 541 TGTGATGCTTATATCGTTCCTCATGAGATCATTAACCATTTACATTCCTTAAGAAATGAGTT 600
Db 558 TGTGATGCTTATATCGTTCCTCATGAGATCATTAACCATTTACATTCCTTAAGAAATGAGTT 617
QY 601 ATCAGCTAGCGAGTTGGCTGCTGAGAGCCCTTCTATCTGGTCGAGGAATCTGTCAAA 660
Db 618 ATCAGCTAGCGAGTTGGCTGCTGAGAGCCCTTCTATCTGGTCGAGGAATCTGTCAAA 677
QY 661 TTCAAGAACCTATCGCGACAAAATAGCGATAACACTTCAAGAAACAACTGGGTACCTTC 720
Db 678 TTCAAGAACCTATCGCGACAAAATAGCGATAACACTTCAAGAAACAACTGGGTACCTTC 737
QY 721 TGTAAAGCAATCCAGGAACCTAATACTAAACAAGCAACAAGCAACAACTAACAGTCA 780
Db 738 TGTAAAGCAATCCAGGAACCTAATACTAACAAGCAACAAGCAACAACTAACAGTCA 797
QY 781 AGCAAGTCAAGTAANTGACATTCGATGCTCTTGAACAGCTCTACAACTGCTTTGAG 840
Db 798 AGCAAGTCAAGTAANTGACATTCGATGCTCTTGAACAGCTCTACAACTGCTTTGAG 857
QY 841 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGCACAATCAAGTCGAAC 900
Db 858 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGCACAATCAAGTCGAAC 917
QY 901 AGCTAGAGGTTGTGAGTGGCCACGAGAGATCAATTAACCTTCAATCCCTTACTCTCAAT 960
Db 918 AGCTAGAGGTTGTGAGTGGCCACGAGAGATCAATTAACCTTCAATCCCTTACTCTCAAT 977
QY 961 GTCTGAATTTGGAACAGGATCGCTGTATTTATTCCTTCTGTTATCGTTTCAACCAATTC 1020
Db 978 GTCTGAATTTGGAACAGGATCGCTGTATTTATTCCTTCTGTTATCGTTTCAACCAATTC 1037
QY 1021 GGTACCAAGATTTCAAGGCGAGAACCAACCAAGTCCCAACCGGACTCCGGAACCTAGTCCAGG 1080
Db 1038 GGTACCAAGATTTCAAGGCGAGAACCAACCAAGTCCCAACCGGACTCCGGAACCTAGTCCAGG 1097
QY 1081 CCGCAACCTGCGACCAAAATCTTAAATAGACTCAAATTTCTTTGTTAGTACGCTGGT 1140
```


780 TGTAAAGCAATCCGAGAACTACAAATATCTAACCAAGCAACACAGCAACACTAACAGTCA 839
781 AGCAAGTCAAAAGTAAATGACATTTGATAGTCTCTTGAACAGCTCTACAAACTGCCCTTTGAG 840
840 AGCAAGTCAAAAGTAAATGACATTTGATAGTCTCTTGAACAGCTCTACAAACTGCCCTTTGAG 899
841 TCAACGACATGTAGAACTGATGCTGCTCTTTGATCCAGCAAAATCAAACTGCAAC 900
900 TCAACGACATGTAGAACTGATGCTGCTCTTTGATCCAGCAAAATCAAACTGCAAC 959
901 AGCTAGAGGTGCTGAGTGCACACGAGATCATTTACCACTTTCACTCTCTCAAAAT 960
960 AGCTAGAGGTGCTGAGTGCACACGAGATCATTTACCACTTTCACTCTCTCAAAAT 1019
961 GTCTGAATTTGAAGAAAGCAATGCTGCTGATTTATTTCCCTCTGTTTCAAAACCAATG 1020
1020 GTCTGAATTTGAAGAAAGCAATGCTGCTGATTTATTTCCCTCTGTTTCAAAACCAATG 1079
1021 GTTACAGATTCAGGCGCAGAACCAAGTGCACACCGACTCGGAACCTAGTCCAGG 1080
1080 GTTACAGATTCAGGCGCAGAACCAAGTGCACACCGACTCGGAACCTAGTCCAGG 1139
1081 CCCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTTGTTGTTAGTCACTGGT 1140
1140 CCCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTTGTTGTTAGTCACTGGT 1199
1141 ACGAAAGTTGGGGAAGATATGATTTGCAAGAAAGGCAATCTCTGTTATGCTTTTC 1200
1200 ACGAAAGTTGGGGAAGATATGATTTGCAAGAAAGGCAATCTCTGTTATGCTTTTC 1259
1201 GAAAGATTTACCATCTGAACTGTTAAATCTTTGAAAGCAAGTATCAAAACAGAGAG 1260
1260 GAAAGATTTACCATCTGAACTGTTAAATCTTTGAAAGCAAGTATCAAAACAGAGAG 1319
1261 TGTTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1320
1320 TGTTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1379
1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATAAGGTCG 1380
1380 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATAAGGTCG 1439
1381 TAATCTGATTTCCAGCTTAGCAAAATTTAGACGCTTGAATGATGATGATGATGATGAT 1440
1440 TAATCTGATTTCCAGCTTAGCAAAATTTAGACGCTTGAATGATGATGATGATGATGATGAT 1499
1441 TAAAGAAAAATTTGATGATGATTTATTTGGCAATCTTAGCAACCAATTTACCATCCAGAGCG 1500
1500 TAAAGAAAAATTTGATGATGATTTATTTGGCAATCTTAGCAACCAATTTACCATCCAGAGCG 1559
1501 ACTTGGCAACCAATTTCTCAATTTAGTATGATGATGATGATGATGATGATGATGATGATGAT 1560
1560 ACTTGGCAACCAATTTCTCAATTTAGTATGATGATGATGATGATGATGATGATGATGATGAT 1619
1561 AGCTGATAAGTATCAAGCTCAGATGTTTACATTTTGGTGAACATGATATATATATATATAT 1620
1620 AGCTGATAAGTATCAAGCTCAGATGTTTACATTTTGGTGAACATGATATATATATATATATAT 1679
1621 TGAAGGAGATGCATATGATTAAGCCCTCATATGAGGCAATGATGATGATGATGATGATGATGAT 1680
1680 TGAAGGAGATGCATATGATTAAGCCCTCATATGAGGCAATGATGATGATGATGATGATGATGAT 1739
1691 CCTTCTGATAGGAAAGTTGAGCTCAAGCTCATATTAAGAAAGAAAGTATCTTACC 1740
1740 CCTTCTGATAGGAAAGTTGAGCTCAAGCTCATATTAAGAAAGAAAGTATCTTACC 1799
1741 TCCATCTCCAGACGAGATGTTAAGCAAAATCCCACTGGAGATAGTGCAGAGCTATTTA 1800
1800 TCCATCTCCAGACGAGATGTTAAGCAAAATCCCACTGGAGATAGTGCAGAGCTATTTA 1859
1801 CAATCGTGTGAAGGGGAAAAACGAATTCCTGCTCGTCACTTCCATATATGTTGTGAGCA 1860

1860 CAATCGTGTGAAGGGGAAAAACGAATTCCTGCTCGTTCGACTCCATATATGTTGTGAGCA 1919
1861 TACAGTTGAGGTTAAAAACCGTAAATTTGATATTTCTCTATAAGGATCAATTTACCAATATAT 1920
1920 TACAGTTGAGGTTAAAAACCGTAAATTTGATATTTCTCTATAAGGATCAATTTACCAATATAT 1979
1921 TAAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
1980 TAAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 2039
1981 TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCTCTGACGAAGCTCCACATTTCAATGA 2040
2040 TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCTCTGACGAAGCTCCACATTTCAATGA 2099
2041 TGGATGGGCAATGCTCAGTGAAGCATGTTGTTAGCAAGAAAGACACAGTGAAGATCCAAA 2100
2100 TGGATGGGCAATGCTCAGTGAAGCATGTTGTTAGCAAGAAAGACACAGTGAAGATCCAAA 2159
2101 TAAAGACTTCAAGGCGGATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2160
2160 TAAAGACTTCAAGGCGGATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2219
2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGG 2220
2220 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGG 2279
2221 AGTAACGGATTTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGTTTACGAATAA 2280
2280 AGTAACGGATTTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGTTTACGAATAA 2339
2281 TTTGACTCTTCAATATGATTAACATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
2340 TTTGACTCTTCAATATGATTAACATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2399
2341 GTTGTAAAAAGGAGTAATCTTTCATCTGTAAGTAAGGAAAAATAAAC 2389
2400 GTTGTAAAAAGGAGTAATCTTTCATCTGTAAGTAAGGAAAAATAAAC 2448

RESULT 6
AF291695 2541 bp DNA linear BCT 15-MAR-2001
LOCUS
DEFINITION Streptococcus pneumoniae pneumococcal histidine triad A protein
ACCESSION AF291695
VERSION AF291695.1 GI:13345012
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 2541)
AUTHORS Wizenmann,T.M., Heinrichs,J.H., Adamou,J.E., Erwin,A.L., Kunsch,C.,
Choi,G.H., Barash,S., Rosen,C.A., Masure,H.R., Tuomanen,E.,
Gayle,A., Brewah,Y.A., Walsh,W., Barren,P., Lathigra,R., Hanson,M.,
Langermann,S., Johnson,S. and Koenig,S.
TITLE Use of a whole genome approach to identify vaccine molecules
affording protection against Streptococcus pneumoniae infection
JOURNAL Infect. Immun. 69 (3), 1593-1598 (2001)
MEDLINE 21116976
PUBMED 11179332
REFERENCE 2 (bases 1 to 2541)
AUTHORS Choi,G.H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Molecular Biology, Human Genome Sciences,
Inc., 9410 Key West Ave., Rockville, MD 20850, USA
FEATURES
source Location/Qualifiers
1..2541
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/strain="N4"
/db_xref="taxon:1313"
91..2541
gene


```

Db      2161  TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAGAAAGCAGAGAGTTTGTTCGGAA 2220
Qy      2221  AGTAACCGGATTCAGTCTGAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
Db      2221  AGTAACCGGATTCAGTCTGAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280
Qy      2281  TTTGACTCTTCAAATTTATGGATTAACAATAGTATCATGCGCAGAGCAGAAATAATCTTGC 2340
Db      2281  TTTGACTCTTCAAATTTATGGATTAACAATAGTATCATGCGCAGAGCAGAAATAATCTTGC 2340
Qy      2341  GTTCTTAAAGGAGTAATCTTCCATCTGTAAGTAAGGAAATAATAAC 2389
Db      2341  GTTCTTAAAGGAGTAATCTTCCATCTGTAAGTAAGGAAATAATAAC 2389

RESULT 4
LOCUS   BD268048                2451 bp    DNA        linear    PAT 17-JUL-2003
DEFINITION Streptococcus pneumoniae protein and immunogenic fragments for
vaccines.
ACCESSION BD268048
VERSION   BD268048.1 GI:33077816
KEYWORDS JP 2002532561-A/6.
SOURCE   Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
          Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
          Streptococcus.
REFERENCE 1 (bases 1 to 2451)
AUTHORS   Johnson,L.S., Koenig,S. and Adamou,J.E.
TITLE      Streptococcus pneumoniae protein and immunogenic fragments for
JOURNAL   Patent: JP 2002532561-A 6 02-OCT-2002;
COMMENT   MEDIMUNE INC
          OS Streptococcus pneumoniae
          PN JP 2002532561-A/6
          PD 02-OCT-2002
          PF 21-DEC-1999 JP 2000589215
          PR 21-DEC-1998 US 60/113048
          PI LESLIE S JOHNSON,SCOTT KOENIG,JOHN E ADAMOU
          PC A61K39/09,A61K38/00,A61P31/10,A61K37/02
          CC Streptococcus pneumoniae protein and immunogenic fragments for
          vaccines
          CC Key Location/Qualifiers
          FH source 1..2451
          FT /organism='Streptococcus pneumoniae'.
          FT location/Qualifiers
          1..2451
          /organism='Streptococcus pneumoniae'
          /mol_type='genomic DNA'
          /db_xref='taxon:1313'

FEATURES
source
Query Match 100.0%; Score 2389; DB 6; Length 2451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 TTTCTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 60
Db      60 TTTCTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 119
Qy      61 TATAGATGGAACCAAGCGCAGCAAAACCGAGAAATTTGACTCTCGATGAGTTAGCAA 120
Db      120 TATAGATGGAACCAAGCGCAGCAAAACCGAGAAATTTGACTCTCGATGAGTTAGCAA 179
Qy      121 GCGTGAAGGAATCAATGCTGACCAATCGTCATCAAGATAACAGACCAAGGCTATGTAC 180
Db      180 GCGTGAAGGAATCAATGCTGACCAATCGTCATCAAGATAACAGACCAAGGCTATGTAC 239
Qy      181 TTCACATGGCGACCACTATCATTTATTAATGGTAAGGTTCTTTAGCGCTATCATCAG 240
Db      240 TTCACATGGCGACCACTATCATTTATTAATGGTAAGGTTCTTTAGCGCTATCATCAG 299

```

```

Qy      241 TGAAGAAATTACTCATGAAAGATCCAACTATAGCTAAAGATGAGGATATTCTTAATGA 300
Db      300 TGAAGAAATTACTCATGAAAGATCCAACTATAGCTAAAGATGAGGATATTCTTAATGA 359
Qy      301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAGGATGC 360
Db      360 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAGGATGC 419
Qy      361 TGGCCACGGGGATAACGTCGGTACAAAGAGGAAATCAATCGACAAATAACAAGAGCATAG 420
Db      420 TGGCCACGGGGATAACGTCGGTACAAAGAGGAAATCAATCGACAAATAACAAGAGCATAG 479
Qy      421 TCAACATCGTGAAGGTGGAACCTCCAAAGAAACGATGTGCTGTGTCCTTGGCAGCTTCGCA 480
Db      480 TCAACATCGTGAAGGTGGAACCTCCAAAGAAACGATGTGCTGTGTCCTTGGCAGCTTCGCA 539
Qy      481 AGGACGCTATPACACAGATGATGGTTATATCTTTTAATGCTTCTGATATCATAGAGATAC 540
Db      540 AGGACGCTATPACACAGATGATGGTTATATCTTTTAATGCTTCTGATATCATAGAGATAC 599
Qy      541 TGGTGATGCTTATATCGTTCCTCATCGAGATCAATACCATTACATTCCTTAAGATGAGTT 600
Db      600 TGGTGATGCTTATATCGTTCCTCATCGAGATCAATACCATTACATTCCTTAAGATGAGTT 659
Qy      601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCCCTATCTGTGTCGAGGAAATCTGTCAA 660
Db      660 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCCCTATCTGTGTCGAGGAAATCTGTCAA 719
Qy      661 TTCAAGAACCTATCGCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
Db      720 TTCAAGAACCTATCGCGCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 779
Qy      721 TGTAAAGCAATCCAGGACCTACAAATCTACAGCAAGCAACAGCAAGCAACTAACAGTCA 780
Db      780 TGTAAAGCAATCCAGGAACTACAAATCTACAGCAAGCAACAGCAAGCAACTAACAGTCA 839
Qy      781 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTTGAAAACAGCTCTACAAACTGCTTTGAG 840
Db      840 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTTGAAAACAGCTCTACAAACTGCTTTGAG 899
Qy      841 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGCACAATCAAGTCCGCAAC 900
Db      900 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGCACAATCAAGTCCGCAAC 959
Qy      901 AGCTAGAGGTGTTGCGAGTGCACACGAGATCATTTACCACTTCATCCCTTACTCTCAAT 960
Db      960 AGCTAGAGGTGTTGCGAGTGCACACGAGATCATTTACCACTTCATCCCTTACTCTCAAT 1019
Qy      961 GTCTGAATTTGGAAGACGAATCGCTCGTATATTATTTCCCTTTCGTTTATCGTTCAAAACCATG 1020
Db      1020 GTCTGAATTTGGAAGACGAATCGCTCGTATATTATTTCCCTTTCGTTTATCGTTCAAAACCATG 1079
Qy      1021 GGTACACAGATTCAAGGCGCAGAACCAACAGTGCACACCGACTCCGGAACCTAGTCCAGG 1080
Db      1080 GGTACACAGATTCAAGGCGCAGAACCAACAGTGCACACCGACTCCGGAACCTAGTCCAGG 1139
Qy      1081 CCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGCTGGT 1140
Db      1140 CCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGCTGGT 1199
Qy      1141 ACGAAAAGTTGGGGAAGGATATGTTATTCGAAGAAAAGGGGATCTCTCGTTATGCTTTTC 1200
Db      1200 ACGAAAAGTTGGGGAAGGATATGTTATTCGAAGAAAAGGGGATCTCTCGTTATGCTTTTC 1259
Qy      1201 GAAAGATTTACCATCTGAAACTGTTAAAATCTTTGAAAGCAAGTTATCAAAACAAGAGAG 1260
Db      1260 GAAAGATTTACCATCTGAAACTGTTAAAATCTTTGAAAGCAAGTTATCAAAACAAGAGAG 1319
Qy      1261 TGTTCACACACTTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
Db      1320 TGTTCACACACTTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1379
Qy      1321 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGSCCTTGTGTGNAATAAAGGTCG 1380

```


QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAAAGGTTAAGGAAATAATCGTGTTCCTTA 60
 Db 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAAAGGTTAAGGAAATAATCGTGTTCCTTA 60
 QY 61 TATAGATGGAAGAAACAAAGCGACGCAAAACAAAGCGAGAAATTTGACTCTGATGAGGTTAGCAA 120
 Db 61 TATAGATGGAAGAAACAAAGCGACGCAAAACAAAGCGAGAAATTTGACTCTGATGAGGTTAGCAA 120
 QY 121 GCCTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCAC 180
 Db 121 GCCTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCAC 180
 QY 181 TTCACATGCGGACCACTATCATTAATTACAAATGGTAAGGTTCCCTTATGACGCTATCATCAG 240
 Db 181 TTCACATGCGGACCACTATCATTAATTACAAATGGTAAGGTTCCCTTATGACGCTATCATCAG 240
 QY 241 TGAAGAAATTAATCATGAAGATCCAACTAATAAGCTAAAGATGAGGATATTGTTAATGA 300
 Db 241 TGAAGAAATTAATCATGAAGATCCAACTAATAAGCTAAAGATGAGGATATTGTTAATGA 300
 QY 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGC 360
 Db 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGC 360
 QY 361 TGCCCAACGGGATAACGTCGTTACAAAGAGGAAATCAATCGACAAACAAAGAGCATAG 420
 Db 361 TGCCCAACGGGATAACGTCGTTACAAAGAGGAAATCAATCGACAAACAAAGAGCATAG 420
 QY 421 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGGCTTGGCAGCTTCGCA 480
 Db 421 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGGCTTGGCAGCTTCGCA 480
 QY 481 AGGACGCTATACACAGATGATGTTATATCTTTAATGTTCTGATATCATGAGGATAC 540
 Db 481 AGGACGCTATACACAGATGATGTTATATCTTTAATGTTCTGATATCATGAGGATAC 540
 QY 541 TGGTATGCTTATATCGTTTCTCATGAGATCATTAACCATTAATTCCTTAAGATGAGTT 600
 Db 541 TGGTATGCTTATATCGTTTCTCATGAGATCATTAACCATTAATTCCTTAAGATGAGTT 600
 QY 601 ATCAGCTAGCGATTTGGCTGCTGCAGAACGCTTCCTATCTGCTGAGGAAATCTGTCAAA 660
 Db 601 ATCAGCTAGCGATTTGGCTGCTGCAGAACGCTTCCTATCTGCTGAGGAAATCTGTCAAA 660
 QY 661 TTCAAGAACTTATCGCCGACAAATAAGCATTAACACTTCAAGAAACAACTGGGTACCTTC 720
 Db 661 TTCAAGAACTTATCGCCGACAAATAAGCATTAACACTTCAAGAAACAACTGGGTACCTTC 720
 QY 721 TGTAAGCAATCCAGGAATACAAATACTAACAACAGCAACAAACAGCACTAACAGTCA 780
 Db 721 TGTAAGCAATCCAGGAATACAAATACTAACAACAGCAACAAACAGCACTAACAGTCA 780
 QY 781 AGCAAGTCAAAGTATGATGATGCTTCTTGAACACAACTGCAAGCACTGCTTGAAG 840
 Db 781 AGCAAGTCAAAGTATGATGATGCTTCTTGAACACAACTGCAAGCACTGCTTGAAG 840
 QY 841 TCAACGACATGATGATGCTTCTTGAACACAACTGCAAGCACTGCTTGAAG 900
 Db 841 TCAACGACATGATGATGCTTCTTGAACACAACTGCAAGCACTGCTTGAAG 900
 QY 901 AGCTAGAGTGTTCAGTGCACACGAGATCATTAACCACTTATCCCTTACTCTCAAT 960
 Db 901 AGCTAGAGTGTTCAGTGCACACGAGATCATTAACCACTTATCCCTTACTCTCAAT 960
 QY 961 GTCCTGAATTTGAAGAACGATCGCTGATTAATTCCTTCTGTTATCGTTCAACCATTTG 1020
 Db 961 GTCCTGAATTTGAAGAACGATCGCTGATTAATTCCTTCTGTTATCGTTCAACCATTTG 1020
 QY 1021 GGTACCAATTTCAAGCCAGAACAAACCAAGTCCAAACCGACTCCGGAACCTAGTCCAGG 1080
 Db 1021 GGTACCAATTTCAAGCCAGAACAAACCAAGTCCAAACCGACTCCGGAACCTAGTCCAGG 1080

QY 1081 CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGTGGT 1140
 Db 1081 CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGTGGT 1140
 QY 1141 ACAGAAAGTTGGGAAAGGATATGTTATTCGAAGAAAGGGCATCTCTGTTATGCTTTTC 1200
 Db 1141 ACAGAAAGTTGGGAAAGGATATGTTATTCGAAGAAAGGGCATCTCTGTTATGCTTTTC 1200
 QY 1201 GAAGAAATTTACCAATCTGTAAGCTGTTAAATCTTGAAGCAAGTTATCAAAACAAAGAG 1260
 Db 1201 GAAGAAATTTACCAATCTGTAAGCTGTTAAATCTTGAAGCAAGTTATCAAAACAAAGAG 1260
 QY 1261 TGTTCACACATCTTAACTGCTTAAAGAAAGAAATGTTGCTCCTCGTACCAAGATTTTA 1320
 Db 1261 TGTTCACACATCTTAACTGCTTAAAGAAAGAAATGTTGCTCCTCGTACCAAGATTTTA 1320
 QY 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTTGNAATTAAGGCTG 1380
 Db 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTTGNAATTAAGGCTG 1380
 QY 1381 TAATTCGTATTTCCAGCCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA 1440
 Db 1381 TAATTCGTATTTCCAGCCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA 1440
 QY 1441 TAAAGAAATTTGCTAGATGATTTATTTGGCAATCTTAGACCAATTAACCATCCAGAGCG 1500
 Db 1441 TAAAGAAATTTGCTAGATGATTTATTTGGCAATCTTAGACCAATTAACCATCCAGAGCG 1500
 QY 1501 ACTTGGCAACCAAAATCTCAATTCGATATATCTGAAGACGAAGTTCTGATTTCTCAAT 1560
 Db 1501 ACTTGGCAACCAAAATCTCAATTCGATATATCTGAAGACGAAGTTCTGATTTCTCAAT 1560
 QY 1561 AGCTGATAAGTATACAACTGCTAGATGTTTACATTTTGTGATGACATGATATCATGTA 1620
 Db 1561 AGCTGATAAGTATACAACTGCTAGATGTTTACATTTTGTGATGACATGATATCATGTA 1620
 QY 1621 TGAAGAGATGCTATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
 Db 1621 TGAAGAGATGCTATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680
 QY 1681 CTTTCTGATTAAGGAAAGTTGAGCTCAAGCTTATCTAAAGAAAGGTTATCTTACC 1740
 Db 1681 CTTTCTGATTAAGGAAAGTTGAGCTCAAGCTTATCTAAAGAAAGGTTATCTTACC 1740
 QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGCAGCAGCTTATTA 1800
 Db 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGCAGCAGCTTATTA 1800
 QY 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA 1860
 Db 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA 1860
 QY 1861 TACAGTTGAGGTTAAACCGGTAAATTTGATTAATTCCTCAAGGATCATTAACATATAT 1920
 Db 1861 TACAGTTGAGGTTAAACCGGTAAATTTGATTAATTCCTCAAGGATCATTAACATATAT 1920
 QY 1921 TAAATTTGCTGTTTGTGATGATCACAATCAAGCTCCAAATGGTATACCTTGAAGA 1980
 Db 1921 TAAATTTGCTGTTTGTGATGATCACAATCAAGCTCCAAATGGTATACCTTGAAGA 1980
 QY 1981 TTTGTTGCGAGTAAAGTACTAGTAGAACCCCTGACGAACCTGCCACATTTCTAATGA 2040
 Db 1981 TTTGTTGCGAGTAAAGTACTAGTAGAACCCCTGACGAACCTGCCACATTTCTAATGA 2040
 QY 2041 TGGATGGGCAATGTCAGTGAAGCATGTTTAGGCAAGAAACCAACAGTGAAGATCCAAA 2100
 Db 2041 TGGATGGGCAATGTCAGTGAAGCATGTTTAGGCAAGAAACCAACAGTGAAGATCCAAA 2100
 QY 2101 TAAAGAACTTCAAAGCGGATGAAGCCAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC 2160
 Db 2101 TAAAGAACTTCAAAGCGGATGAAGCCAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC 2160
 QY 2161 TCAAGTAGAGCTGAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTTCCTTTCGAA 2220


```
QY 781 ACBAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAACTCGCTTTGAG 840
Db 781 ACBAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAACTCGCTTTGAG 840
QY 841 TCAACGACATGTAGAACTCTGATGGCTTCTCTTTCATCCAGACAAATCACAAAGTCGAAC 900
Db 841 TCAACGACATGTAGAACTCTGATGGCTTCTCTTTCATCCAGACAAATCACAAAGTCGAAC 900
QY 901 AGCTAGAGTGTTCAGTCCACAGGAGATCATTTACCACTTCATCCCTTACTCTCAAT 960
Db 901 AGCTAGAGTGTTCAGTCCACAGGAGATCATTTACCACTTCATCCCTTACTCTCAAT 960
QY 961 GTCCTAATTTGGAAGACGAATCGCTCGTATTATTCCTTTCGTTATCGTTTCAACCAATG 1020
Db 961 GTCCTAATTTGGAAGACGAATCGCTCGTATTATTCCTTTCGTTATCGTTTCAACCAATG 1020
QY 1021 GGTACCAAGTTCAAGGCCAGAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1080
Db 1021 GGTACCAAGTTCAAGGCCAGAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1080
QY 1081 CCCGAACTCGACCAAAATCTTAAATAGACTCAAAATCTCTTTCGTTAGTCAGTGTG 1140
Db 1081 CCCGAACTCGACCAAAATCTTAAATAGACTCAAAATCTCTTTCGTTAGTCAGTGTG 1140
QY 1141 ACGAAAGTTGGGAAGGATATGTTTCGAAGAAAGGCGCATCTCTCGTTATGTTCTTGC 1200
Db 1141 ACGAAAGTTGGGAAGGATATGTTTCGAAGAAAGGCGCATCTCTCGTTATGTTCTTGC 1200
QY 1201 GAAAGATTACCATCTGAAACTGTTAAAGTCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
Db 1201 GAAAGATTACCATCTGAAACTGTTAAAGTCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
QY 1261 TGTTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
Db 1261 TGTTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
QY 1321 TGATAAGCATATACTCTGTAAGTCTGATGCTCATAAAGCTGTTTGNAAATAAGGTCG 1380
Db 1321 TGATAAGCATATACTCTGTAAGTCTGATGCTCATAAAGCTGTTTGNAAATAAGGTCG 1380
QY 1381 TAATTCGTGATTTCCAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA 1440
Db 1381 TAATTCGTGATTTCCAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA 1440
QY 1441 TAAAGAAATTTGGTAGATGATTTATGSCATTCCTAGCACCAATTTACCATCCAGAGCG 1500
Db 1441 TAAAGAAATTTGGTAGATGATTTATGSCATTCCTAGCACCAATTTACCATCCAGAGCG 1500
QY 1501 ACTTGGCAACCAAAATCTCAAAATTCAGTATGATGAGCAAGGTTGCTGATTCCTCAATT 1560
Db 1501 ACTTGGCAACCAAAATCTCAAAATTCAGTATGATGAGCAAGGTTGCTGATTCCTCAATT 1560
QY 1561 AGCTGATAAGTATCAACGCTCAGATGATGATTTTGTATGATGATGATGATGATGATGATG 1620
Db 1561 AGCTGATAAGTATCAACGCTCAGATGATGATTTTGTATGATGATGATGATGATGATGATG 1620
QY 1621 TGAAGGAGTGCATATGTAACGCTCATATGGCCATAGTCTGATGATGATGATGATGATGATG 1680
Db 1621 TGAAGGAGTGCATATGTAACGCTCATATGGCCATAGTCTGATGATGATGATGATGATGATG 1680
QY 1681 CTTTCTGATAGGAAAGGTTGACGCTCAAGCTTACTTAAAGAAAGGTTATCTTACC 1740
Db 1681 CTTTCTGATAGGAAAGGTTGACGCTCAAGCTTACTTAAAGAAAGGTTATCTTACC 1740
QY 1741 TCCATCTCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATGATGACAGCTATTATTA 1800
Db 1741 TCCATCTCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATGATGACAGCTATTATTA 1800
QY 1801 CAATCGTGTGAAGGGGAAAGAAAGCAATTCACCTGCTTCCACTTCCATATATGTTGAGCA 1860
Db 1801 CAATCGTGTGAAGGGGAAAGAAAGCAATTCACCTGCTTCCACTTCCATATATGTTGAGCA 1860
```

```
QY 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTTCTCATAGGATCATTAACCAATAT 1920
Db 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTTCTCATAGGATCATTAACCAATAT 1920
QY 1921 TAAATTTGCTTTGGTTTGTATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
Db 1921 TAAATTTGCTTTGGTTTGTATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
QY 1981 TTTGTTTCGGACGATTAAGTACTAGTAGAACACACCTGACGAACTCCACATTTCTAATGA 2040
Db 1981 TTTGTTTCGGACGATTAAGTACTAGTAGAACACACCTGACGAACTCCACATTTCTAATGA 2040
QY 2041 TCGATGGGCGAATGCCAGTGAAGCATGTGTAGGCAAGAAAGCACACAGTGAAGATCCAAA 2100
Db 2041 TCGATGGGCGAATGCCAGTGAAGCATGTGTAGGCAAGAAAGCACACAGTGAAGATCCAAA 2100
QY 2101 TAAAGAACTTCAAGCGGATGAAGAGCCAGTGAAGAAACACCTGCTGAGCCAGAAAGTCCC 2160
Db 2101 TAAAGAACTTCAAGCGGATGAAGAGCCAGTGAAGAAACACCTGCTGAGCCAGAAAGTCCC 2160
QY 2161 TCAAGTAGAGACTGAAAGAGTAGAGCCCACTCAAGAGCGAAGATTTGCTTGGCGAA 2220
Db 2161 TCAAGTAGAGACTGAAAGAGTAGAGCCCACTCAAGAGCGAAGATTTGCTTGGCGAA 2220
QY 2221 AGTAACGGAATTCAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
Db 2221 AGTAACGGAATTCAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
QY 2281 TTTGACTCTTCAAAATTTAGGATAACAATAGTATCATGCGAAGCAGAAAAATTTACTTGC 2340
Db 2281 TTTGACTCTTCAAAATTTAGGATAACAATAGTATCATGCGAAGCAGAAAAATTTACTTGC 2340
QY 2341 GTTGTAAAGGAAGTAATCTTCTATCTGTAAGTGAAGAAATAAATAAC 2389
Db 2341 GTTGTAAAGGAAGTAATCTTCTATCTGTAAGTGAAGAAATAAATAAC 2389

RESULT 3
BD063274
LOCUS BD063274 2389 bp DNA linear PAT 27-AUG-2002
DEFINITION Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063274
VERSION BD063274.1 GI:22608877
KEYWORDS JP 2001505415-A/28.
SOURCE unidentified
ORGANISM Streptococcus pneumoniae antigens and vaccines
REFERENCE 1 (bases 1 to 2389)
AUTHORS Kunsch,C.A., Choi,G.H., Johnson,S.L. and Hromocky,j,A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: JP 2001505415-A 28 24-APR-2001;
HUMAN GENOME SCIENCES INC
COMMENT FN JP 2001505415-A/28
PD 24-APR-2001
PR 30-OCT-1997 JP 1998520667
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH,GIL H CHOI,SYDNOR L JOHNSON,ALEX HROMOCKYJ PC
C12N15/31,C12N5/18,C12N1/21,C07K14/315,C12Q1/68,A61K39/09, PC
G01N33/569,
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1..2389
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

Query Match 100.0%; Score 2389; DB 6; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

| | | | |
|----------|------|---|--------|
| QY | 1381 | TAATTCGTGATTTCCAAAGCCTTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA | 1440 |
| Db | 1381 | TAATTCGTGATTTCCAAAGCCTTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA | 1440 |
| QY | 1441 | TAAAGAAAATTTGGTAGATGATTTATTGGCATTTCTTAGACACCAATTTACCCATCCAGAGCG | 1500 |
| Db | 1441 | TAAAGAAAATTTGGTAGATGATTTATTGGCATTTCTTAGACACCAATTTACCCATCCAGAGCG | 1500 |
| QY | 1501 | ACTTTGGCAAAACCAAAATTCCTCAAAATTGAGTATFACCTGAAGACGAAAGTTTCGTATTGCTCAATT | 1560 |
| Db | 1501 | ACTTTGGCAAAACCAAAATTCCTCAAAATTGAGTATFACCTGAAGACGAAAGTTTCGTATTGCTCAATT | 1560 |
| QY | 1561 | AGCTGATAAGTATACAAACGTCAGATGGTTTACATTTTGTATGAACATGATATAATCAGTGA | 1620 |
| Db | 1561 | AGCTGATAAGTATACAAACGTCAGATGGTTTACATTTTGTATGAACATGATATAATCAGTGA | 1620 |
| QY | 1621 | TGAAGGAGATGCATATGTAAACGCTCATATGGGCCATAGTCTACTGGATTGGAAAAAGATAG | 1680 |
| Db | 1621 | TGAAGGAGATGCATATGTAAACGCTCATATGGGCCATAGTCTACTGGATTGGAAAAAGATAG | 1680 |
| QY | 1681 | CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATFACCTAAAGAAAAAGTATCCTPACC | 1740 |
| Db | 1681 | CCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATFACCTAAAGAAAAAGTATCCTPACC | 1740 |
| QY | 1741 | TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACCTGGAGATAGTGCAGCAGCTATTTA | 1800 |
| Db | 1741 | TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACCTGGAGATAGTGCAGCAGCTATTTA | 1800 |
| QY | 1801 | CAATTCGTGTGAAGGGGAAAAAGAAATTCACCTCGCTTGCATTCCTCATATATGGTTGAGCA | 1860 |
| Db | 1801 | CAATTCGTGTGAAGGGGAAAAAGAAATTCACCTCGCTTGCATTCCTCATATATGGTTGAGCA | 1860 |
| QY | 1861 | TACAGTTGAGGTTAAAAAACGGTAATTTTGATTTATTCCTCATAGGATCAATTTACCATAATAT | 1920 |
| Db | 1861 | TACAGTTGAGGTTAAAAAACGGTAATTTTGATTTATTCCTCATAGGATCAATTTACCATAATAT | 1920 |
| QY | 1921 | TAAATTTGCTTTGGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA | 1980 |
| Db | 1921 | TAAATTTGCTTTGGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA | 1980 |
| QY | 1981 | TTTGTWTGCGACGATTAAGTACTACGTPAGAACCCCTGACGAACGTCACATTTCTAATGA | 2040 |
| Db | 1981 | TTTGTWTGCGACGATTAAGTACTACGTPAGAACCCCTGACGAACGTCACATTTCTAATGA | 2040 |
| QY | 2041 | TGGATGGGCAATGCCAGTGCAGATGTTTAGCCAGAAGAGACACACAGTGAAGTCCAAA | 2100 |
| Db | 2041 | TGGATGGGCAATGCCAGTGCAGATGTTTAGCCAGAAGAGACACACAGTGAAGTCCAAA | 2100 |
| QY | 2101 | TAAAGAACTTCAAAGCGGATGAAGACCGAGTAGAGGAAAAACCTGCTGAGCCAGAAAGTCCC | 2160 |
| Db | 2101 | TAAAGAACTTCAAAGCGGATGAAGACCGAGTAGAGGAAAAACCTGCTGAGCCAGAAAGTCCC | 2160 |
| QY | 2161 | TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAGACGAGAAGTTTGTCTGCGAA | 2220 |
| Db | 2161 | TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAGACGAGAAGTTTGTCTGCGAA | 2220 |
| QY | 2221 | AGTAAACGGATTCTAGTCTGAAAGCCTATGCAACAGAAACTCTAGCTGGTTTACGAAATAA | 2280 |
| Db | 2221 | AGTAAACGGATTCTAGTCTGAAAGCCTATGCAACAGAAACTCTAGCTGGTTTACGAAATAA | 2280 |
| QY | 2281 | TTTGACTCTTCAAATTTATGATAACATAGTATCATGCGCAAGACGAGAAAAATTTACTTGC | 2340 |
| Db | 2281 | TTTGACTCTTCAAATTTATGATAACATAGTATCATGCGCAAGACGAGAAAAATTTACTTGC | 2340 |
| QY | 2341 | GTGTGTTAAAAAGGAGTAATCTTTCATCTGTAAGTAAAGGAAAAATAAAC | 2399 |
| Db | 2341 | GTGTGTTAAAAAGGAGTAATCTTTCATCTGTAAGTAAAGGAAAAATAAAC | 2399 |
| RESULT 2 | | | |
| AR340956 | | | |
| LOCUS | | | |
| | | 2389 bp | DNA |
| | | AR340956 | linear |
| | | PAT 17-AUG-2003 | |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 22:17:38 ; Search time 9275 Seconds
(without alignments)
11164.044 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTACAGTTGGACTGT.....TAAGTAAGGAAAAATAAAC 2389

Scoring table:

OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database :

GenEmbl:

1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rtd:*
36: em.htg.mam:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|-----------|---------------------|
| 1 | 2389 | 100.0 | 2389 | 6 | AR120265 | AR120265 Sequence |
| 2 | 2389 | 100.0 | 2389 | 6 | AR340956 | AR340956 Sequence |
| 3 | 2389 | 100.0 | 2389 | 6 | BD063274 | BD063274 Streptoco |
| 4 | 2389 | 100.0 | 2451 | 6 | BD268048 | BD268048 Streptoco |
| 5 | 2389 | 100.0 | 2451 | 6 | AR344447 | AR344447 Sequence |
| 6 | 2389 | 100.0 | 2541 | 1 | AF291695 | AF291695 Streptoco |
| 7 | 2338 | 97.9 | 2406 | 6 | AX569139 | AX569139 Sequence |
| 8 | 2338 | 97.9 | 8195 | 6 | AR218862 | AR218862 Sequence |
| 9 | 2338 | 97.9 | 8195 | 6 | BD003774 | BD003774 Polynucle |
| C 10 | 2338 | 97.9 | 10256 | 1 | AE007418 | AE007418 Streptoco |
| C 11 | 2338 | 97.9 | 349980 | 6 | AX571763 | AX571763 Sequence |
| C 12 | 1624 | 68.0 | 10320 | 1 | AE008479 | AE008479 Streptoco |
| 13 | 927 | 38.8 | 20035 | 2 | SPNEU1915 | SPNEU1915 Streptoco |
| 14 | 326 | 13.6 | 702 | 2 | SPNEU1929 | SPNEU1929 Streptoco |
| 15 | 200 | 8.4 | 2535 | 1 | AF340221 | AF340221 Streptoco |
| 16 | 148 | 6.2 | 2523 | 6 | BD263585 | BD263585 Novel str |
| 17 | 148 | 6.2 | 2523 | 6 | AX343072 | AX343072 Sequence |
| 18 | 148 | 6.2 | 2647 | 6 | BD263590 | BD263590 Novel str |
| 19 | 148 | 6.2 | 2647 | 6 | AX343073 | AX343073 Sequence |
| 20 | 53 | 2.2 | 504 | 6 | BD229970 | BD229970 Human com |
| C 21 | 53 | 2.2 | 504 | 6 | BD229971 | BD229971 Human com |
| 22 | 53 | 2.2 | 2166 | 12 | AF340222 | AF340222 Synthetic |
| 23 | 53 | 2.2 | 2290 | 6 | AR120270 | AR120270 Sequence |
| 24 | 53 | 2.2 | 2290 | 6 | AR340961 | AR340961 Sequence |
| 25 | 53 | 2.2 | 2290 | 6 | BD063279 | BD063279 Streptoco |
| 26 | 53 | 2.2 | 2359 | 6 | AR219011 | AR219011 Sequence |
| 27 | 53 | 2.2 | 2359 | 6 | BD003923 | BD003923 Polynucle |
| 28 | 53 | 2.2 | 2457 | 1 | AF318954 | AF318954 Streptoco |
| 29 | 53 | 2.2 | 2457 | 6 | AX569137 | AX569137 Sequence |
| 30 | 53 | 2.2 | 2478 | 6 | BD229972 | BD229972 Human com |
| 31 | 53 | 2.2 | 2481 | 6 | BD264091 | BD264091 Streptoco |
| 32 | 53 | 2.2 | 2517 | 1 | AF318955 | AF318955 Streptoco |
| 33 | 53 | 2.2 | 2517 | 6 | AX568778 | AX568778 Sequence |
| 34 | 53 | 2.2 | 2531 | 6 | BD268046 | BD268046 Streptoco |
| 35 | 53 | 2.2 | 2531 | 6 | BD268049 | BD268049 Streptoco |
| 36 | 53 | 2.2 | 2531 | 6 | AR344445 | AR344445 Sequence |
| 37 | 53 | 2.2 | 2531 | 6 | AR344448 | AR344448 Sequence |
| 38 | 53 | 2.2 | 2639 | 6 | BD263591 | BD263591 Novel str |
| 39 | 53 | 2.2 | 2639 | 6 | AX343074 | AX343074 Sequence |
| 40 | 53 | 2.2 | 11931 | 1 | AE007403 | AE007403 Streptoco |
| 41 | 53 | 2.2 | 12372 | 1 | AE008464 | AE008464 Streptoco |
| 42 | 53 | 2.2 | 75874 | 2 | SPNEU1907 | SPNEU1907 Streptoco |
| 43 | 53 | 2.2 | 232807 | 2 | SPNEU1901 | SPNEU1901 Streptoco |
| 44 | 53 | 2.2 | 349980 | 6 | AX571762 | AX571762 Sequence |
| 45 | 50 | 2.1 | 492 | 12 | AF340223 | AF340223 Synthetic |
| 46 | 44 | 1.8 | 973 | 6 | AR219123 | AR219123 Sequence |
| 47 | 44 | 1.8 | 973 | 6 | BD004035 | BD004035 Polynucle |
| 48 | 38 | 1.6 | 3171 | 6 | BD263631 | BD263631 Novel str |
| 49 | 30 | 1.3 | 40 | 6 | AR120406 | AR120406 Sequence |
| 50 | 30 | 1.3 | 40 | 6 | BD268045 | BD268045 Streptoco |
| C 51 | 30 | 1.3 | 40 | 6 | AR341097 | AR341097 Sequence |
| C 52 | 30 | 1.3 | 40 | 6 | AR344444 | AR344444 Sequence |
| C 53 | 30 | 1.3 | 40 | 6 | BD063415 | BD063415 Streptoco |
| 54 | 27 | 1.1 | 36 | 6 | BD268043 | BD268043 Streptoco |
| 55 | 27 | 1.1 | 36 | 6 | AR344442 | AR344442 Sequence |
| 56 | 27 | 1.1 | 37 | 6 | AR120405 | AR120405 Sequence |
| 57 | 27 | 1.1 | 37 | 6 | AR341096 | AR341096 Sequence |
| 58 | 27 | 1.1 | 37 | 6 | BD063414 | BD063414 Streptoco |
| 59 | 26 | 1.1 | 35 | 6 | BD268044 | BD268044 Streptoco |
| 60 | 26 | 1.1 | 35 | 6 | AR344443 | AR344443 Sequence |
| 61 | 25 | 1.0 | 33 | 6 | BD263613 | BD263613 Novel str |
| 62 | 25 | 1.0 | 33 | 6 | AX343114 | AX343114 Sequence |
| 63 | 25 | 1.0 | 34 | 6 | BD263615 | BD263615 Novel str |
| 64 | 25 | 1.0 | 34 | 6 | AX343116 | AX343116 Sequence |
| 65 | 24 | 1.0 | 1342 | 6 | AR120328 | AR120328 Sequence |

| | | | |
|-----|----|---|-----|
| 562 | QY | SKEKVAAQAQYTKEGILPPSPDADVKANPTGDSAAAIYNRVKGEKRIPILVRPYMVEHT | 621 |
| | | : : : : : : : : : : : : : : : : : : : : | |
| 302 | Db | SEAEAAAQAAYAKEGLTPPSTDHQSNGTEAKGAEALYNRVKAAKKVPLDRFYNLYQT | 361 |
| | | : : : : : : : : : : : : : : : : : : : : | |
| 622 | QY | VEVKGNLLIIPHKDHYHNIKFAWPDHNYKAPNGYTLEDLFATIKYYVEHPDERPHSNDG | 681 |
| | | : : : : : : : : : : : : : : : : : : : : | |
| 362 | Db | VEVKNGSLIIPHVDHYHNIKFEWFDEGYEAPKGYTTLELATVKKYYVEHENRPHSDNG | 421 |
| | | : : : : : : : : : : : : : : : : : : : : | |
| 682 | QY | WGNASEHLVGKKHSEDPNKNFKAD-----EEPVV | 711 |
| | | : : : : : : : : : : : : : : : : : : : : | |
| 422 | Db | FGNASDVQRNKGQADNTQETKPSEEKQPOTEKPEETPREEKPOSEPSPKTEEPPE | 481 |
| | | : : : : : : : : : : : : : : : : : : : : | |
| 712 | QY | ETPASPEVPQVETEKVBEAQLKAEEVLIAKVTDSSLKANATETLAGLRNNLTLOIMDNNSI | 771 |
| | | : : : : : : : : : : : : : : : : : : : : | |
| 482 | Db | ESPESESEFPQVETEKVBEELKREADELGIQODPIIKSNAKETITGLKNLLFTGTQNNTI | 541 |
| | | : : : : : : : : : : : : : : : : : : : : | |
| 772 | QY | MAEAKLLALLKGS-----NPFSSVS | 791 |
| | | : : : : : : : : : : : : : : : : : : : : | |
| 542 | Db | MAEAEKLIALLESKGPNNOPSOLS | 566 |

RESULT 25

```

US-10-324-143-26
; Sequence 26, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 555
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-26

```

| | | | | |
|-----------------------|---|---------------------|------------|-------------|
| Query Match | 45.3%; | Score 1888; | DB 15; | Length 555; |
| Best Local Similarity | 65.7%; | Pred. No. 1.8e-127; | | |
| Matches 364; | Conservative 59; | Mismatches 97; | Indels 34; | Gaps 3; |
| 266 | DIISLLKQLYKLPISORHVESDGLVPDPAQITSTRTARGVAVPHGDHYHFIPYSQMSLEE 325 | | | |
| db | 1 DIISLLKQLYKLPISORHVESDGLIFDPAQITSTRTARGVAVPHGNHYHFIPYEQMSLEK 60 | | | |
| 266 | RIARIPLRVSRNHWVDSRPPSPQPTPESSPQFPAPNLK-IDSN---SSIVSOLVR 381 | | | |
| db | 61 RIARIPLRVSRNHWVDSRPPSPQPTPESSPQFPAPNPQAPSPNIDKLVKCAVR 120 | | | |
| 382 | KVGEYGVFEKGISRVYFAKDLPSETVKNLESKLQESVSHLTAKKENVAPRQDFYD 441 | | | |
| db | 121 KVGEYGVFEENGVSRYIPAKNLSAETAAGIDSKLQESLSHKLGAKKTDLPSSDRFYN 180 | | | |
| 442 | KAYNLLTEAHKALFYXKGRNSQFQALDKLLERLNDSTNKEKLVDDLLAFAPITPHERL 501 | | | |
| db | 181 KAYDLLARHQDILLDNKGRQVDFEALDNLRLKQVSSDKVKLVDDLLAFAPITPHERL 240 | | | |
| 502 | GKENSQLEYTEDEVRVTAQLADKYTTSDBGYIFDEHDIISDEGDVYVTPMHGSHHWIKOSL 561 | | | |
| db | 241 GKPNQATYTDDBIQVAKLAGKYTTEDEGYIFDPDRITISDEGDVYVTPMTHSHWIKOSL 300 | | | |

| | | |
|----|-----|---|
| Qy | 682 | WGNASEHVGLGKHSDENPNKFKAD-----EEPYE 711 |
| Dd | 422 | FGNASDHVQRNKGQADTNQTEKPESEKQPOTEKPEETPREKPQSEKPSKPTEPSEE 481 |
| Qy | 712 | ETPAEPVPQVETEKEVAQLKEAEVLAKVTDSLSKANATETLAGLRNNLTQIQMDNNSI 771 |
| Dd | 482 | ESPEESEBPQVETEKEVEKLREAADILGKIQDPIIKSNAKETITGLKNLLFGTQDNNTI 541 |
| Qy | 772 | MAEAELLALLKGSNPSSVSKEKI 795 |
| Dd | 542 | MAEAELLALLKBSKGPDLTEEQI 565 |

```

RESULT 23
US-09-884-465A-382
; Sequence 382, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charlier, Nathalie
; APPLICANT: Cuellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 382
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or no
; NAME/KEY: MISC FEATURE
; LOCATION: (557)..(557)
; OTHER INFORMATION: Xaa = Glycine or nothi
; NAME/KEY: MISC FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: Xaa = Proline or nothi
; US-09-884-465A-382

```

| | | | | |
|---------------------------|--------|---|------------|--------------|
| Query Match | 45.4%; | Score 1891.5; | DB 10; | Length 1365; |
| Best Local Similarity | 65.0%; | Pred. No. 4e-127; | | |
| Matches 367; Conservative | 60; | Mismatches 99; | Indels 39; | Gaps 4; |
| | | | | |
| QY | 266 | DIDSLLLKQLYKLP LLSQRHVESDGLVFDEPAQTTSRTARGVAVPHGDHYHFIPIYSOMSELEE | 325 | |
| Db | 2 | DIDSLLLKQLYKLP LLSQRHVESDGLVDFDAQITTSRTARGVAVPHGNYHHFIPYEOMSELEK | 61 | |
| | | ::::::::::::::::::::: | | |
| QY | 326 | RIARIITPLRYRSHNVWPDSPRPQPOTPPESPQGQPAPNLK-IDSN---SSAVSQSLVR | 381 | |
| | | ::::::::::::::::::::: | | |
| Db | 62 | RIARIITPLRYRSHNVWPDSPRPQPOTPPESPQPAPNPQPAPNPIDEKLVEAR | 121 | |
| | | ::::::::::::::::::::: | | |
| QY | 382 | KVGEGVVFEKGISRYVFAKDLPSTVTNKLESKLSKQBSVSHTLTAKKNVAPRDQRYD | 441 | |
| | | ::::::::::::::::::::: | | |
| Db | 122 | KVGDGVVFEENGVSRYI PAKNLSAETAAIGIDSKLAKQESLSHLGAKKTDLTPSSDRFFYN | 181 | |
| | | ::::::::::::::::::::: | | |
| QY | 442 | KAYNILTEAHKALFXNKGNSDFOALDKLLERLNDESTNKEKLVDDLLAF LAPI THPERL | 501 | |
| | | ::::::::::::::::::::: | | |
| Db | 182 | KAYDLLARIHODLLDNKGQVDPEALDNLRLKUVSSDKVLVDLLAF LAPI THPERL | 241 | |
| | | ::::::::::::::::::::: | | |
| QY | 502 | GKFNQSIEYTEDEVRIAQLADKYTTSDGYIFDEHDIIISDEGDAYVTPMHGHSHWKKSOL | 561 | |
| | | ::::::::::::::::::::: | | |
| Db | 242 | GKPNAOITYTDDBIQVAKLAGKYTTBDEGYIFOPRDIITSDEGDAYVTPMTHGHSHWKKSOL | 301 | |
| | | ::::::::::::::::::::: | | |

| | | | |
|----|-----|--|-----|
| Qy | 562 | SDKEKVAQAQYKTEKGIILPPSPDADVKANPTCDSSAAIYNRVKGBKRIPLVLRPLPYMVBEHT | 622 |
| Db | 302 | SEASRAAAQAQYAKSEKGLTPSPDHOOSGNTRAKGAEAIYNRVKAAKKVPLDRMPYLNQYT | 361 |
| Qy | 622 | VEVKNGNLIIPKHDIYHNHKEAFDHDYTKAPNGYTTLEDLFATIKYVVEHPDERPHSNDG | 681 |
| Db | 362 | VEVKNGSLIIPHYDIYHNHKEWFDEGLYEAPKGYTTLEDLLATVKYVVEHPNERPHSNDG | 421 |
| Qy | 682 | WGNASEHVLGKHDSHEDNNKNFKAD-----PEPVE | 711 |
| Db | 422 | FGNASDHVORNKNGQADNTQTEKPSSEKPOTEKPBETPREEEKPOSEKPSKPTPEEPEE | 481 |
| Qy | 712 | ETPABPEPQVETEKVEACLAKEAEVLLAKVTDSSILKANATETLAGLRNNLLTIQIMDNNSI | 771 |
| Db | 482 | ESPESEEPQVETEKVEEKLREAEDLLGKIQDPIIKSNAKETLTGLKNNLLFTGTQDNNTI | 541 |
| Qy | 772 | MAEAEKLLALLKGS-----NPSSVS | 791 |
| Db | 542 | MAEAEKLLALLKESXXXNNQMPSQLS | 566 |

```

RESULT 24
US-10-324-143-53
; Sequence 53, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELLETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 53
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified streptococcus pneumoniae protein sequence
US-10-324-143-53

```

| | | | | | | | |
|-----------------------|--------|----------------|-------------------------------------|--------------------|------------|--------|-------|
| Query Match | 45.4%; | Score | 1891.5; | DB | 15; | Length | 1365; |
| Best Local Similarity | 65.0%; | Pred. No. | 4e-127; | | | | |
| Matches | 367; | Conservative | 60; | Mismatches | 99; | Indels | 39; |
| Gaps | 4 | | | | | | |
| Qy | 266 | DIDSLLKQLYKLP | LQSQRHVESDGLVFDEPAQITSTRTARGVAVPHG | DHYHFIPYSQMSLEEE | 325 | | |
| | | | | | | | |
| Db | 2 | DIDSLLKQLYKLP | LQSQRHVESDGLVFDEPAQITSTRTARGVAVPHG | NHNYHFIPYEQMSLEK | 61 | | |
| | | | | | | | |
| Qy | 326 | RIARIIPLYRNHNV | PDSPRPQSPQPTPEPSPGQPAPNLK- | IDSN--- | SSLVSQLVQR | 381 | |
| | | | | | | | |
| Db | 62 | RIARIIPLYRNHNV | PDSPRPQSPQPTPEPSPGQPAPNPQAPSPID | EKLKVEAVR | 121 | | |
| | | | | | | | |
| Qy | 382 | KVGEGVVFEKGI | SRVFAKDLFSETVTKNLESKLQKESVSHTLTAK | KENVAPRQEFVD | 441 | | |
| | | | | | | | |
| Db | 122 | KVGDGVVFEENG | SVRYIPAKNLSAETAAIGDSKLAKQESLHKLAKKT | DTLDPSSDREFVN | 181 | | |
| | | | | | | | |
| Qy | 442 | KAYNMLTEAHKAL | FYNKGNSSFOALDKLLERLNDGESTNKEKLV | DDLLAFIAPITHPERL | 501 | | |
| | | | | | | | |
| Db | 182 | KAYDILLARHQD | LNDKNGQVDFEALDNLERLKVSSDKVL | VDDLLAFIAPIRPERL | 241 | | |
| | | | | | | | |
| Qy | 502 | GKDNSCIYTED | EVRIAQLADKYTTSDGYIFDEHDIISDEG | DAYVTPHMGHSHWIKDGL | 561 | | |
| | | | | | | | |
| Db | 242 | GKFNQAITVTD | DIIOVAKLAGKYTTEDGYIFOPRDIITSDEG | DAYVTPHMGHSHWIKDGL | 301 | | |
| | | | | | | | |

```
QY 774 EAEKLLALLKGS 785
Db 601 EAEKLLALLKES 612

RESULT 21
US-09-884-465A-383
; Sequence 383, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 383
; LENGTH: 1126
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC FEATURE
; LOCATION: (557)..(557)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-383

Query Match 45.6%; Score 1899; DB 10; Length 1126;
Best Local Similarity 64.9%; Pred. No. 8.5e-128;
Matches 366; Conservative 63; Mismatches 101; Indels 34; Gaps 3;

QY 266 DIDLKQLYKLPLSQRHVESDGLVDFPAQITSTRTARGVAVPHGHHYHFIPIYQMSLEEE 325
Db 2 DIDLKQLYKLPLSQRHVESDGLVDFPAQITSTRTARGVAVPHGHHYHFIPIYQMSLEEK 61

QY 326 RIARIIPLYRSNHWVPSRPEQSPQPTPEPSPQAPNPK-IDSN--SSLVSQLVR 381
Db 62 RIARIIPLYRSNHWVPSRPEQSPQPTPEPSPQAPNPK-IDSN--SSLVSQLVR 121

QY 382 KVGEGYVFEKIGIRYVFAKDLPSKESKLSKQESVSHLTAKKENVAPRDQEFYD 441
Db 122 KVGEGYVFEENGVSRYIPAKNLSAETAAGIDSKLAKQESLSHKLGAKTDLPPSSDRFYN 181

QY 442 KAYNLLTEAHKALFNKGRNSDFQALDKLERLNDESNKEKLVDDLLAFAPITHPERL 501
Db 182 KAYDILLARIHQDLLDNKGRQVDFEALDNLLERLKDVSDDKVLVDDIILAFAPIRHPERL 241

QY 502 GKPNISOIYETDEVRIAQLADKYTTSDGYIFDEHDIISDEGDVYVTPMHGSHWIGKDSL 561
Db 242 GKPNAIITYTDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDVYVTPMTHSHWIKKDSL 301

QY 562 SDKEKVAQAAYTKKGIILPPSPADVKANPTGDSAAAIYNRVKGEKRIPLVRLPYMVHET 621
Db 302 SEAEERAAQAAYAKEKGLTPPSTDHDSNGTEAKGAEALYNRVKAARKVPLDRMPYLNQYT 361

QY 622 VEVKNGNLIIPHKHQHYHNKFAWFDHTYKAPNGYTTLEDLPAITKYVVEHPDERPHSNDG 681
Db 362 VEVKNGSLIIPHYDHYHNKFEWFDGLYKAPGYTTLEDLLATVKYVVEHPNERPHSDNG 421
```

```
QY 582 WGNASEHVLGKKDHSDEPNKNFKAD-----ERPVE 711
Db 422 FGNASDHVQRNKGQADTNQTEKSEKPKQTEKEETPREKQSEKSPKTEPREE 481

QY 712 ETPAEPEVPQVETKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTFLQIMDNNSI 771
Db 482 ESPESEEPQVETKVEKLEAREADLLGKIQDPIIKSNAKETTLGLKXNLLFGTQDNNTI 541

QY 772 MAEAEKLLALLKGSNPSVSKEKI 795
Db 542 MAEAEKLLALLKESKXXDLTBEQI 565

RESULT 22
US-10-324-143-54
; Sequence 54, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54
; LENGTH: 1126
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-54

Query Match 45.6%; Score 1898; DB 15; Length 1126;
Best Local Similarity 64.9%; Pred. No. 1e-127;
Matches 366; Conservative 63; Mismatches 101; Indels 34; Gaps 3;

QY 266 DIDLKQLYKLPLSQRHVESDGLVDFPAQITSTRTARGVAVPHGHHYHFIPIYQMSLEEE 325
Db 2 DIDLKQLYKLPLSQRHVESDGLVDFPAQITSTRTARGVAVPHGHHYHFIPIYQMSLEEK 61

QY 326 RIARIIPLYRSNHWVPSRPEQSPQPTPEPSPQAPNPK-IDSN--SSLVSQLVR 381
Db 62 RIARIIPLYRSNHWVPSRPEQSPQPTPEPSPQAPNPK-IDSN--SSLVSQLVR 121

QY 382 KVGEGYVFEKIGIRYVFAKDLPSKESKLSKQESVSHLTAKKENVAPRDQEFYD 441
Db 122 KVGEGYVFEENGVSRYIPAKNLSAETAAGIDSKLAKQESLSHKLGAKTDLPPSSDRFYN 181

QY 442 KAYNLLTEAHKALFNKGRNSDFQALDKLERLNDESNKEKLVDDLLAFAPITHPERL 501
Db 182 KAYDILLARIHQDLLDNKGRQVDFEALDNLLERLKDVSDDKVLVDDIILAFAPIRHPERL 241

QY 502 GKPNISOIYETDEVRIAQLADKYTTSDGYIFDEHDIISDEGDVYVTPMHGSHWIGKDSL 561
Db 242 GKPNAIITYTDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDVYVTPMTHSHWIKKDSL 301

QY 562 SDKEKVAQAAYTKKGIILPPSPADVKANPTGDSAAAIYNRVKGEKRIPLVRLPYMVHET 621
Db 302 SEAEERAAQAAYAKEKGLTPPSTDHDSNGTEAKGAEALYNRVKAARKVPLDRMPYLNQYT 361

QY 622 VEVKNGNLIIPHKHQHYHNKFAWFDHTYKAPNGYTTLEDLPAITKYVVEHPDERPHSNDG 681
Db 362 VEVKNGSLIIPHYDHYHNKFEWFDGLYKAPGYTTLEDLLATVKYVVEHPNERPHSDNG 421
```

```

; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272-66

Query Match 63.6%; Score 2649.5; DB 9; Length 763;
Best Local Similarity 66.8%; Pred. No. 3.1e-182;
Matches 519; Conservative 77; Mismatches 132; Indels 49; Gaps 8;

QY 1 SYELGHYQRTV-KENNRVSYIDGQATQKTNLTPEVSKREGINAEQIVIKITDQYV 59
DB 2 SYELGRHQGVKESNRVSYIDGQAGQKAENLTPEVSKREGINAEQIVIKITDQYV 61
QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGYYVVKVDGYVYLKD 119
DB 62 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDSIDVNEIKGGYVVKVNGYYVYLKD 121
QY 120 AAHADNVRTKEINRQKQESQHRGGTTPRNDGAVALARSGRYTTDDGYIFNASDIID 179
DB 122 AAHADNIRTKETIKRQKQESHNNHNS---RADNAVAARAQGRYTTDDGYIFNASDIID 178
QY 180 TGDYAVPHGDHYHYTPKNELSAELAAAFISGRGNLSNRTYRRQNSDTSRINWVP 239
DB 179 TGDYAVPHGDHYHYTPKNELSAELAAAEAYWNG-----KQGRSPSSSSYNA 227
QY 240 SVSNPGTNTNTNNSNTNSQAQSDNDIDSLLKQYKLPISQRHVESDGLVFPQAQITSR 299
DB 228 NPAQPLSENHNLTVTPTHQ--NQGENISLLRELVAKPLSEHRVSDGLIFPQAQITSR 286
QY 300 TARGVAVPHGDHYHYFIPYQMSLEBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSP 359
DB 287 TARGVAVPHGNHYHYFIPYQMSLEKRIARIIPLYRSNHWVPDSRPEQSPQPTPEPSP 346
QY 360 GPQAPNLK-IDSN---SSLVSQLVKVGEGYFEEKISRYVFAKDLPSRTVKNLESKL 415
DB 347 SPOFAPNPQAPSNPIDKLVKAVRVKVGEGYFEEGVSYIPAKDLSAETAAGIDSKL 406
QY 416 SKQESVSHLTAKKENVAPRDOBFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLN 475
DB 407 AKQESLSHKGAKTDLPSDRFYNAKAYDILLARIHQDLNDKGRQVDFEALNDLLERLK 466
QY 476 DESTNKEKLVDDLLAFAPITHPERLGKNSQIETTEDVIRIAQLADKYTTSDGYIFDEH 535
DB 467 DVKSDKVLXVOLLAFAPIRHPERLGKNAQITTYTDDETQVAKLAGKYTTEDGYIFDPR 526
QY 536 DIISDEGDYAVTHMGHSHWIKGDSISDKKVAQAQYATKEGILPSPDADKANPTGDS 595
DB 527 DITSDEGDYAVTHMTHSHWIKKDSLSAEARAAQAYAKEGLTPPSTHDQDSGNTAKG 586
QY 596 AAAYNRVKGKRIPLVRIPYVMEHTVEVKNGNLIIPKDHYNHNIKFAWDDHYKAPNG 655
DB 587 AEALYNRVKAARKVPLDRPYNLYQIVVEKNGSLIIPHYDHYHNIKFEWDEGLYAPKG 646
QY 656 YTTLEDLPATIKYVVEHPDPRHNDGKNASEHVLGKKHSDSDPNKNFKADEE-----P 709
DB 647 YTTLEDLATYKYVVEHPNERPHSDNGFGNASDHVQRNKGQADTNQTEKPESEKPKTEK 706
QY 710 VEET-----PAPPEVPOVETEKVEAQLEAEVLLAKVTD 743
DB 707 EETPREEKQSEKPSKPTPEEPSESESESESESESESESESESESESESESESESESESE 763

```

RESULT 20

```

US-10-324-143-22
; Sequence 22, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; PRIOR FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-22

```

```

Query Match 52.1%; Score 2170; DB 15; Length 613;
Best Local Similarity 68.5%; Pred. No. 9.2e-148;
Matches 419; Conservative 59; Mismatches 100; Indels 34; Gaps 3;

QY 208 AEAFISGRGNLSNRTYRRQNSDTSRINWVPVSPGTCITNTNNSNTNSQAQSDNDI 267
DB 1 AEAFISGRGNLSNRTYRRQNSDTSRINWVPVSPGTCITNTNNSNTNSQAQSDNDI 60
QY 268 DSSLKQYKLPISQRHVESDGLVFPQAQITSRARGVAVPHGDHYHYFIPYQMSLEBERI 327
DB 61 DSSLKQYKLPISQRHVESDGLVFPQAQITSRARGVAVPHGDHYHYFIPYQMSLEBERI 120
QY 328 ARIIPLYRSNHWVPDSRPEQSPQPTPEPSPQAPNLK-IDSN---SSLVSQLVKRV 383
DB 121 ARIIPLYRSNHWVPDSRPEQSPQPTPEPSPQAPNLK-IDSN---SSLVSQLVKRV 180
QY 384 GEGYVFEKIGISRYVFAKDLPSRTVKNLESKLSQSVSHLTAKKENVAPRDOBFYDKA 443
DB 181 GDGYVFEENGVSRYIPAKNLSAETAAGIDSKLAKQESLSHKLGAKTDLPSDRFYNA 240
QY 444 YNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTNKEKLVDDLLAFAPITHPERLGK 503
DB 241 YDLARIHQDLNDKGRQVDFEALNDLLERLKDVSDKVKLVDDLLAFAPIRHPERLGK 300
QY 504 PMSQIETTEDVIRIAQLADKYTTSDGYIFDDEHDIISDEGDYAVTHMGHSHWIKGDSLSD 563
DB 301 PNAQITTYTDDETQVAKLAGKYTTEDGYIFDPRDITSDEGDYAVTHMTHSHWIKGDSLSE 360
QY 564 KEKVAQAQYATKEGILPSPDADKANPTGDSAAAYNRVKGKRIPLVRLPYMEHTVE 623
DB 361 AEARAAQAYAKEGLTPPSTHDQDSGNTAKGAEALYNRVKAARKVPLDRMPNLQYVTE 420
QY 624 VKNGNLIIPKDHYNHNIKFAWDDHYKAPNGYTTLEDLPATIKYVVEHPDPRHSDNGWG 683
DB 421 VKNGSLIIPHYDHYHNIKFEWDEGLYAPKGKYTTLEDLATYKYVVEHPNERPHSDNGFG 480
QY 684 NASEHVLGKKHSDSDPNKNFKADEE-----EETVEET 713
DB 481 NASDHVQRNKGQADTNQTEKPESEKPKTEKPEETPREEKQSEKPSKPTPEEPSEES 540
QY 714 PAPEVPOVETEKVEAQLEAEVLLAKVTDSSLSKANATETLAGLRNLTITQIDNNNSIMA 773
DB 541 PESEEPQVETEKVEEKLREADLLGKIQDPIIKSNAKETLTGLKNNLFTGTQDNNITMA 600

```


Db 366 SLOPAPNPQAPNSPID--EKLKVEAVKVGDDGVFFENGVSRYIPAKDLISAETAGIDS 423
Qy 414 KLSQESVSHLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLER 473
Db 424 KLAQESLSHKLGAKTDLPSDDREFYNKAYDILLARHQDLNDKGRQVDFEVLNLLER 483
Qy 474 LNDESTNKEKLVDDLLAFAPITHPERLGRPNQSIETEDVRIQALADKVTTSDDGYIED 533
Db 484 LKDVSSDKVLVDLILAFAPIRHPERLGRPNQSIETEDVRIQALADKVTTSDDGYIED 543
Qy 534 EHDIIISDEGDAYVTPHMGSHWIKOSLSDEKVAQAQYTKKEGILPPSPDADVKANPTG 593
Db 544 PRDITSDGDAYVTPHMGSHWIKOSLSDEKVAQAQYTKKEGILPPSPDADVKANPTG 603
Qy 594 DSAAIYNRVKGKRIPLVRLPYMVEHTVEYKNGNLIIPHKKDHYHNKIFAWFDDHTYKAP 653
Db 604 KGAELIYNRVKAACKVPLDRMPYNLQYTVKNGSLIIPHYDHYHNKIFEFWDFGLYEAP 663
Qy 654 NGYLEDLFATIKYVVEHPDRPHSDNGWGNASHVGLK-----KHSE----- 697
Db 664 KGSLEDDLATVKYVVEHPDRPHSDNGWGNASHVGLK-----KHSE----- 723
Qy 698 -----DPNKNFKADPEPVEETPAEPEVPQVETEKVEAQLKEAEV 736
Db 724 HPESDEKENHAGLNPSADNLYKPSDTEETEEBAEDTTDEAEIQVENSVINAKIADAEA 783
Qy 737 LKAVTSSSLKANATETLAGLNLLTQIMDNNSIMAEKLLALLKGSNPSV 790
Db 784 LLEKVTDPISIRQNAWETITGLKSSLLTGKONNTISAEDVSLALLKESQPAPI 837
RESULT 18
US-10-324-143-9
; Sequence 9, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-9
Query Match 66.0%; Score 2750.5; DB 15; Length 838;
Best Local Similarity 64.5%; Pred. No. 1.8e-189;
Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;
Qy 1 SYELGLYQARTV-KENNRVSYIDGKQATOKNTLTPDEVSKREGINAEQIVIKITDQGV 59
Db 21 SYELGRHQAGVKESNRVSYIDGQAGQKAENLTPDEVSKREGINAEQIVIKITDQGV 80
Qy 60 TSHGDHYHYNGKVPYDAIISBELLMKDPNKLKDEDIYNEKGGVIVKDGKYYVLKD 119
Db 81 TSHGDHYHYNGKVPYDAIISBELLMKDPNKLKDEDIYNEKGGVIVKDGKYYVLKD 140
Qy 120 AAHADNVRTKEETNRQKSHQREGTGRNDGCAVALARSQGRYTTDDGYIFNASDIID 179

Db 141 AAHADNVRTKEETNRQKSHQSHNHNS---RADNAVAARAQGRYTTDDGYIFNASDIID 197
Qy 180 TGDAYIVPHGDHYHYIPKNELSASELAFAFSLGRNLSNSETYRQNSDNTSRTNWVP 239
Db 198 TGDAYIVPHGDHYHYIPKNELSASELAFAFSLGRNLSNSETYRQNSDNTSRTNWVP 246
Qy 240 SVSNPGTNTNTNNSNTNSQASQNDIDSLKQLYKPLSORHVESDGLVDFDPAQITSR 299
Db 247 NPQVPRUSENHNLTFTYHQ--NQGENISLLBELYAKPLSERHVESDGLVDFDPAQITSR 305
Qy 300 TARGVAVPHGDHYHYFIPYSQMSLEERARIIPLRVRSNHWVDSRPEQSPQPTPPSP 358
Db 306 TARGVAVPHGNHYHFIPYEQMSLEKRIARIIPLRVRSNHWVDSRPEQSPQPTPPSP 365
Qy 359 -----PQOPAPNLKIDSNSLSYQVRYKVGEGVVEEKGISRYVFAKDLPSETVKLES 413
Db 366 SLOPAPNPQAPNSPID--EKLKVEAVKVGDDGVFFENGVSRYIPAKDLISAETAGIDS 423
Qy 414 KLSQESVSHLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLER 473
Db 424 KLAQESLSHKLGAKTDLPSDDREFYNKAYDILLARHQDLNDKGRQVDFEVLNLLER 483
Qy 474 LNDESTNKEKLVDDLLAFAPITHPERLGRPNQSIETEDVRIQALADKVTTSDDGYIED 533
Db 484 LKDVSSDKVLVDLILAFAPIRHPERLGRPNQSIETEDVRIQALADKVTTSDDGYIED 543
Qy 534 EHDIIISDEGDAYVTPHMGSHWIKOSLSDEKVAQAQYTKKEGILPPSPDADVKANPTG 593
Db 544 PRDITSDGDAYVTPHMGSHWIKOSLSDEKVAQAQYTKKEGILPPSPDADVKANPTG 603
Qy 594 DSAAIYNRVKGKRIPLVRLPYMVEHTVEYKNGNLIIPHKKDHYHNKIFAWFDDHTYKAP 653
Db 604 KGAELIYNRVKAACKVPLDRMPYNLQYTVKNGSLIIPHYDHYHNKIFEFWDFGLYEAP 663
Qy 654 NGYLEDLFATIKYVVEHPDRPHSDNGWGNASHVGLK-----KHSE----- 697
Db 664 KGSLEDDLATVKYVVEHPDRPHSDNGWGNASHVGLK-----KHSE----- 723
Qy 698 -----DPNKNFKADPEPVEETPAEPEVPQVETEKVEAQLKEAEV 736
Db 724 HPESDEKENHAGLNPSADNLYKPSDTEETEEBAEDTTDEAEIQVENSVINAKIADAEA 783
Qy 737 LKAVTSSSLKANATETLAGLNLLTQIMDNNSIMAEKLLALLKGSNPSV 790
Db 784 LLEKVTDPISIRQNAWETITGLKSSLLTGKONNTISAEDVSLALLKESQPAPI 837
RESULT 19
US-09-765-272-66
; Sequence 66, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083

```
RESULT 16
US-10-324-143-32
; Sequence 32, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARMAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELLETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-32

Query Match          66.0%; Score 2750.5; DB 15; Length 819;
Best Local Similarity 64.5%; Pred. No. 1.8e-189;
Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;

QY 1 SYELGLYQARTV-KENNRVSYIDGKQATOKTENLTDPDEVSKREGINAEQIVIKITDQGYV 59
Db 2 SYELGRHQAGQVKESNRVSYIDGQAGQKAENLTDPDEVSKREGINAEQIVIKITDQGYV 61
QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIDVNEIKGGYVVKDGYVYVVKLD 119
Db 62 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIDVNEIKGGYVVKDGYVYVVKLD 121
QY 120 AAHADNVRKKEINRQKQHSQHREGGTPRNDGAVALARSQGYRTTDDGYIFNASDIIED 179
Db 122 AAHADNIRTKKEIKRQKQHSNHNS---RADNAVAARAQGYRTTDDGYIFNASDIIED 178
QY 180 TGDAYIVPHGDHYHYIPKKNLSASELAFAFLSGRGNLSNRTYRRQNSDTSRINWVP 239
Db 179 TGDAYIVPHGDHYHYIPKKNLSASELAFAFLSGRGNLSNRTYRRQNSDTSRINWVP 227
QY 240 SVSNPQTTNTSNNNTSQAQSQNDISLLKQLYKLPQSORHVESDGLVFPDPAQITSR 299
Db 228 NPQPRLSENHNLTVTPTYHQ-NOGENISLLRELYAKPLSERHVESDGLVFPDPAQITSR 286
QY 300 TARGVAVPHGDHYHYIPYQMSLEBERIARIIPLYRSNHNWVDSRPEQSPQPTPEPS- 358
Db 287 TARGVAVPHGNHYHYIPYQMSLEKRIARIIPLYRSNHNWVDSRPEQSPQPTPEPS 346
QY 359 -----PGPOPAPNLKIDSNSLSVLQVRKVGEGYVEEKGISRYVFAKOLPSETVKNLES 413
Db 347 SIQAPNPQAPSNPID--EKLKVEARKVGDGYVEENGVSRYIPAKOLSAETAAGIDS 404
QY 414 KLSQESVSHLTITAKENAVPRDQEFYDKAYNMLTEAHKALFNKGRNDSDFQALDKLLER 473
Db 405 KLAQESLSHLGAKTDLPSDREFFYNKAYDILLARIHQDLNKGQVDFEVLNLLER 464
QY 474 INDESTNKEKLVDDLLAFAPITHPERLQKPNQSQIETEYDEVRIAQLADKYTTSDGYIFD 533
Db 465 LKDVSSDKVLDVDDLLAFAPIRHPERLGKPNQAITYTDDIQAQVAKLAGKYTTEDGYIFD 524
QY 534 EHDIIISDEGDAVVTPHMGHSHWIGKDSLSDEKVAQAQVYTKKGIILPPSPDADVKANPTG 593
Db 525 PRDIISDEGDAVYTPHMTSHHWIKKDSUSEABRAAAQAYAKEKGLTPPTSTDHQDSGNTSEA 584

RESULT 17
US-09-884-465A-8
; Sequence 8, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-8

Query Match          66.0%; Score 2750.5; DB 10; Length 838;
Best Local Similarity 64.5%; Pred. No. 1.8e-189;
Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;

QY 1 SYELGLYQARTV-KENNRVSYIDGKQATOKTENLTDPDEVSKREGINAEQIVIKITDQGYV 59
Db 21 SYELGRHQAGQVKESNRVSYIDGQAGQKAENLTDPDEVSKREGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEIVNEIKGGYVVKDGYVYVVKLD 119
Db 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIDVNEIKGGYVVKDGYVYVVKLD 140
QY 120 AAHADNVRKKEINRQKQHSQHREGGTPRNDGAVALARSQGYRTTDDGYIFNASDIIED 179
Db 141 AAHADNIRTKKEIKRQKQHSNHNS---RADNAVAARAQGYRTTDDGYIFNASDIIED 197
QY 180 TGDAYIVPHGDHYHYIPKKNLSASELAFAFLSGRGNLSNRTYRRQNSDTSRINWVP 239
Db 198 TGDAYIVPHGDHYHYIPKKNLSASELAFAFLSGRGNLSNRTYRRQNSDTSRINWVP 246
QY 240 SVSNPQTTNTSNNNTSQAQSQNDISLLKQLYKLPQSORHVESDGLVFPDPAQITSR 299
Db 247 NPQPRLSENHNLTVTPTYHQ-NOGENISLLRELYAKPLSERHVESDGLVFPDPAQITSR 305
QY 300 TARGVAVPHGDHYHYIPYQMSLEBERIARIIPLYRSNHNWVDSRPEQSPQPTPEPS- 358
Db 306 TARGVAVPHGNHYHYIPYQMSLEKRIARIIPLYRSNHNWVDSRPEQSPQPTPEPS 365
QY 359 -----PGPOPAPNLKIDSNSLSVLQVRKVGEGYVEEKGISRYVFAKOLPSETVKNLES 413
Db 358 TARGVAVPHGNHYHYIPYQMSLEKRIARIIPLYRSNHNWVDSRPEQSPQPTPEPS 365
```

```

Query Match      66.6%; Score 2772; DB 15; Length 838;
Best Local Similarity 65.0%; Pred. No. 5.2e-191;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLVQARTV-KENNRVSYIDGKQATOKTENITPDEVSKREGINAEQIVIKITDQGYV 59
DB 21 SYELGRHQAGQGVKESNRVSYIDGQAGKAEKITPDEVSKREGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEVKGYVIKVDGKYVYLKD 119
DB 81 TSHGDHYYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEIKGYYVIKVDGKYVYLKD 140
QY 120 AAHADNVRTKEEINRQKQEHSHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIID 179
DB 141 AAHADNVRTKEEIKRQKQEHSHHGGGS--NDQAVVAARAQGRYTTDDGYIFNASDIID 198
QY 180 TGDYIVPHGHDYHYIIPYQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 239
DB 199 TGDYIVPHGHDYHYIIPYQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 247
QY 240 SVSNPGTTNTNTNSNTNSQASQNDIDSLLKOLYKPLSORHVESDGLVFDPAQITSR 299
DB 248 NPAQPLSENENLTVTPYHQ--NQCENISSLLRELYAKPLSERHVESDGLIFDPAQITSR 306
QY 300 TARGVAVPHGHDYHYIIPYQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 359
DB 307 TARGVAVPHGHDYHYIIPYQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 366
QY 360 GPQAPNLK-IDSN--SSLSVQLRVKGVGEGVVEEKIGISRYVFAKOLPSETVKNLESKL 415
DB 367 SPQAPNPQAPSNPIDDKLVEAKVRKVGDDGVFEENGVSRYIPAKOLSAETAAGIDSKL 426
QY 416 SKQESVSHLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLN 475
DB 427 AKQESLSHKLGAKKTDLPSSDREFYFNKAYDYLARITHQDLNDKGRQVDFEALDNLLERLK 486
QY 476 DESTNKEKLVDDLLAFAPITHPERLKGPNQSOIETEDVRITAOALADKYTTSDGYIFDEH 535
DB 487 DVPDKVKLVDDLLAFAPIRHPERLKGPNQSOIETEDVRITAOALADKYTTSDGYIFDPR 546
QY 536 DIISDEGDAYVTPHMGHSHWIGKOSLSDEKKEVAAQAYTKEGILPSPDADVKANPTGDS 595
DB 547 DITSDEGDAYVTPHMGHSHWIGKOSLSDEKKEVAAQAYTKEGILPSPDADVKANPTGDS 606
QY 596 AAAIYNRVKGKRIPIPLRVLYPVVHEHTVEKNGNLIIPKHQHYHNIKFAWFDHHTYKAPNG 655
DB 607 ABAYIYNRVKAACKVPLDRMPYLNQYTVVEKNGSLIIPHYDHYHNIKFEWFDGLYEAPKG 666
QY 656 YTLEDLFATIKYVVEHPDERPHSDNGWGNASHVLGK-----KHSE----- 697
DB 667 YTLEDLLATVKYVVEHPDERPHSDNGWGNASHVLGK-----KHSE----- 726
QY 698 -----DPKNFKADEEPEETPAEPEVQVETEKVEAQLKEAEVLL 738
DB 727 ESEKENHAGLNPSADNLKPSDTDETEEBEADTTDEAEIQVENSVINAKIADAEALL 786
QY 739 AKVTSSLLKANATETLAGLNRLNLTQIMDNNSIMAEAEKLLALLKGSNPSSV 790
DB 787 EKVTDPISIRQNAMEITLGLKSSLLGTGKNTTISAEVDSLLALLKESQAPI 838

RESULT 15
US-10-387-783-4
; Sequence 4, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683

```

```

; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-10-387-783-4

Query Match      66.6%; Score 2772; DB 15; Length 838;
Best Local Similarity 65.0%; Pred. No. 5.2e-191;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLVQARTV-KENNRVSYIDGKQATOKTENITPDEVSKREGINAEQIVIKITDQGYV 59
DB 21 SYELGRHQAGQGVKESNRVSYIDGQAGKAEKITPDEVSKREGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEVKGYVIKVDGKYVYLKD 119
DB 81 TSHGDHYYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEIKGYYVIKVDGKYVYLKD 140
QY 120 AAHADNVRTKEEINRQKQEHSHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIID 179
DB 141 AAHADNVRTKEEIKRQKQEHSHHGGGS--NDQAVVAARAQGRYTTDDGYIFNASDIID 198
QY 180 TGDYIVPHGHDYHYIIPYQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 239
DB 199 TGDYIVPHGHDYHYIIPYQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 247
QY 240 SVSNPGTTNTNTNSNTNSQASQNDIDSLLKOLYKPLSORHVESDGLVFDPAQITSR 299
DB 248 NPAQPLSENENLTVTPYHQ--NQCENISSLLRELYAKPLSERHVESDGLIFDPAQITSR 306
QY 300 TARGVAVPHGHDYHYIIPYQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 359
DB 307 TARGVAVPHGHDYHYIIPYQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 366
QY 360 GPQAPNLK-IDSN--SSLSVQLRVKGVGEGVVEEKIGISRYVFAKOLPSETVKNLESKL 415
DB 367 SPQAPNPQAPSNPIDDKLVEAKVRKVGDDGVFEENGVSRYIPAKOLSAETAAGIDSKL 426
QY 416 SKQESVSHLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLN 475
DB 427 AKQESLSHKLGAKKTDLPSSDREFYFNKAYDYLARITHQDLNDKGRQVDFEALDNLLERLK 486
QY 476 DESTNKEKLVDDLLAFAPITHPERLKGPNQSOIETEDVRITAOALADKYTTSDGYIFDEH 535
DB 487 DVPDKVKLVDDLLAFAPIRHPERLKGPNQSOIETEDVRITAOALADKYTTSDGYIFDPR 546
QY 536 DIISDEGDAYVTPHMGHSHWIGKOSLSDEKKEVAAQAYTKEGILPSPDADVKANPTGDS 595
DB 547 DITSDEGDAYVTPHMGHSHWIGKOSLSDEKKEVAAQAYTKEGILPSPDADVKANPTGDS 606
QY 596 AAAIYNRVKGKRIPIPLRVLYPVVHEHTVEKNGNLIIPKHQHYHNIKFAWFDHHTYKAPNG 655
DB 607 ABAYIYNRVKAACKVPLDRMPYLNQYTVVEKNGSLIIPHYDHYHNIKFEWFDGLYEAPKG 666
QY 656 YTLEDLFATIKYVVEHPDERPHSDNGWGNASHVLGK-----KHSE----- 697
DB 667 YTLEDLLATVKYVVEHPDERPHSDNGWGNASHVLGK-----KHSE----- 726
QY 698 -----DPKNFKADEEPEETPAEPEVQVETEKVEAQLKEAEVLL 738
DB 727 ESEKENHAGLNPSADNLKPSDTDETEEBEADTTDEAEIQVENSVINAKIADAEALL 786
QY 739 AKVTSSLLKANATETLAGLNRLNLTQIMDNNSIMAEAEKLLALLKGSNPSSV 790
DB 787 EKVTDPISIRQNAMEITLGLKSSLLGTGKNTTISAEVDSLLALLKESQAPI 838

```

Db 306 TARGVAVPHGHYHFIPIYEQMSELEKRIARIIPLYRSNHWVPOSRPSPQPTPEPSP 365
 QY 360 GPQAPNLKIDNSGLVQLVKYKGVFEEKGISRYVFAKDLPSFVKNLESKLSQOE 419
 Db 366 SPQAPSPNID--GKLVKEAVKVGDDGVFEEGVSRVYIPAKDLISAETAAGIDSKLQOE 423
 QY 420 SVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDEST 479
 Db 424 SLSHKLGTKTDLPSDDREFYNKAYDILARIHQDLDLNGKQVDFEALDNLRLKDVSS 483
 QY 480 NKEKLVDDLLAFAPITHPERLGPNSQIETEDVRIQAOLADKYTTSDGVIYFDEHDIIS 539
 Db 484 DKVXLVEDILAFAPIRHPERLGPNAQITTYDDEIQVAKLAGKYTABDGYIFDPRDITS 543
 QY 540 DEGDAYVTPHMGHSHWIKGDSLSQKVAQAAYTKKGIPLPSPDADVKANPTGDSAAAI 599
 Db 544 DEGDAYVTPHMGHSHWIKGDSLSQKVAQAAYTKKGIPLPSPDADVKANPTGDSAAAI 603
 QY 600 YNRVKGKRIPLVRLPYMVEHTVEVXNGNLIIPHKHYNHNIKFAWFDHHTYKAPNGYTL 659
 Db 604 YNRVKAACKVPLDRMPNLYQTVVEVXNGSLIIPHVDHYNHNIKFEWDEGLYEAAPKGYTL 663
 QY 660 DLFAIKYVVEHDPDRPHSDNGWGNASEHVLGKDHSEDPNNPKADEE-----PVEET 713
 Db 664 DLLATVYVVEHDPDRPHSDNGWGNASEHVLGKDHSEDPNNPKADEE-----PVEET 723
 QY 714 -----PABPEVPOVETEKVEAQLKEAEVLLAKVTTDSLSKANA 750
 Db 724 PREKPOSEKPSKPTEEPSE 783
 QY 751 TETLAGLNNLTQIMDNNSIMAEKLLALLKGS 785
 Db 784 KETLTGLKNNLFTQDNNTIMAEKLLALLKES 818

RESULT 13

US-10-412-862-4
 ; Sequence 4, Application US/10412862
 ; Publication No. US20040052781A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-685
 ; CURRENT APPLICATION NUMBER: US/10/412,862
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 4
 ; LENGTH: 838
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-10-412-862-4

Query Match 66.6%; Score 2772; DB 12; Length 838;
 Best Local Similarity 65.0%; Pred. No. 5.2e-191;
 Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;
 QY 1 SYELGLYQARTV-KENNRYSIDGKATOKENTLTEDVSKRGINAEQIVIKITDQGV 59
 Db 21 STELGHQAGQAKESNRVSYIDGQQAENLTEDVSKRGINAEQIVIKITDQGV 80
 QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYLKDEDIVNEKGGVVIKVDGKYVYLKD 119
 Db 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYLKDEDIVNEKGGVVIKVDGKYVYLKD 140

QY 120 AAHADNVRTKEEINQKOEHSQHREGTPRNDGAVALARSGRYVTTDDGYIFENASDILED 179
 Db 141 AAHADNVRTKEEINQKOEHSQHREGTPRNDGAVALARSGRYVTTDDGYIFENASDILED 198
 QY 180 TGDYIVPHGHYHFIPIYKNELSAELAAAEAFSLGRGNLSRTRYRQNSQNTSRTNWP 239
 Db 199 TGDYIVPHGHYHFIPIYKNELSAELAAAEAFSLGRGNLSRTRYRQNSQNTSRTNWP 247
 QY 240 SVSNPGTNTNTSNNSTNSQASQNDIDSLLKQLYKLPLSRHVHVESGLVDPDAQITSR 299
 Db 248 NPAQPRLSNENLTVTPTYHQ--NOGENISSLLRELYAKPLSERHVHVESGLVDPDAQITSR 306
 QY 300 TARGVAVPHGHYHFIPIYQMSSELEERTARIIPLYRSNHWVPOSRPSPQPTPEPSP 359
 Db 307 TARGVAVPHGHYHFIPIYQMSSELEERTARIIPLYRSNHWVPOSRPSPQPTPEPSP 366
 QY 360 GPQAPNLK-IDSN---SSLVSQLVRKVGEGYVEEKGISRYVFAKDLPSFVKNLESKL 415
 Db 367 SPQAPNPQAPSPNIDPKLVKEAVRKVGDDGVFEEGVSRVYIPAKDLISAETAAGIDSKL 426
 QY 416 SKOESVSHLTAKKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLN 475
 Db 427 AKQESLSHLGAKKXKTDLPSDDREFYNKAYDILARIHQDLDLNGKQVDFEALDNLRLK 486
 QY 476 DESTNKEKLVDDLLAFAPITHPERLGPNSQIETEDVRIQAOLADKYTTSDGVIYFDEH 535
 Db 487 DVPSDKVLVDDLLAFAPITHPERLGPNAQITTYDDEIQVAKLAGKYTTEDGYIFDPR 546
 QY 536 DIISDEGDAYVTPHMGHSHWIKGDSLSQKVAQAAYTKKGIPLPSPDADVKANPTGDS 595
 Db 547 DIISDEGDAYVTPHMGHSHWIKGDSLSQKVAQAAYTKKGIPLPSPDADVKANPTGDS 606
 QY 596 AAALYNRVKGKRIPLVRLPYMVEHTVEVXNGNLIIPHKHYNHNIKFAWFDHHTYKAPNG 655
 Db 607 AAALYNRVKAACKVPLDRMPNLYQTVVEVXNGSLIIPHVDHYNHNIKFEWDEGLYEAAPK 666
 QY 656 YTLEDLFATIKYVVEHDPDRPHSDNGWGNASEHVLGK-----KDHSE----- 697
 Db 667 YTLEDLFATIKYVVEHDPDRPHSDNGWGNASEHVLGK-----KDHSE----- 726
 QY 698 -----DPNKNFKADEPVEETPAPEVPOVETEKVEAQLKEAEVLL 738
 Db 727 ESDEKENHAGLNPSADNLYKPSDTEETEEAEEDTTDEAEIPOVENSVINAKIADAEALL 786
 QY 739 AKYTDLSKANATETLAGLNNLTQIMDNNSIMAEKLLALLKGSNPSV 790
 Db 787 EKVTDPSIRONAMETLTGLKSLLLGTKDNNTISAEVDSLLALLKESQAPAPI 838

RESULT 14

US-10-412-850-4
 ; Sequence 4, Application US/10412850
 ; Publication No. US20040001836A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-686
 ; CURRENT APPLICATION NUMBER: US/10/412,850
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 4
 ; LENGTH: 838
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-10-412-850-4

Db 484 DKVKLVEDILAFIRPHRPERLGKNAQITTTDDBIQVAKLAGKTAEADGVIFDPRDITS 543

QY 540 DEGDAYVTMHGSHWIGKDSLSDEKVAAQAAYTTEKGILPPSPDADYKANPTGDASAAA 599

Db 544 DEGDAYVTHTMTHSWIKDKSLSSEAAAAQAAYAEKGLTPPSTDHQSNGTEAKGAZAI 603

QY 600 YNRVGEKEIPVLRLPYVMVEHTVTVKGNLIIPIHKDHVHNKFAFDFDHTTKAPNGVYLE 659

Db 604 YNRVKAACKVPPLDRMPYNLYTVVEVKGSLIIPHVDHYHNKFEKFDEGLYEAPKGYTLE 663

QY 660 DLPATIKYYVEHPDERPHSDNCGMGNASEHLVGKKDHSBDPNKNFKADE-----PVET 713

Db 664 DLLLATVKKYVEHPNERPHSDNGFNASHDVQRNNKQGADTWQTEKPSEKQTEKPEEET 723

QY 714 -----PAEPVPOVETEKEVAQLKEAEVLLAKVTDSSLKANA 750

Db 724 PREEKPSQSEKPSKPTBEPSESPSESEFPQVETEKVEEKLREAEADLLGKIQDPIKNSA 783

QY 751 TETLAGLRNNLTLOIMDNNSIMAERKLLALLKGS 785

Db 784 KETTLGLKNNLLFGTODNNTIMAEAKLLALLKES 818

RESULT 12
US-10-387-783-10
; Sequence 10, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 3.0
; SEQ ID NO 10
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-10

Query Match 67.0%; Score 2788.5; DB 15; Length 819;
Best Local Similarity 66.7%; Pred. No. 3.2e-192;
Matches 544; Conservative 85; Mismatches 139; Indels 47; Gaps 7;

QY 1 SYELGLYQA-RTKVENNVSVIDGKQATOKTENLTDPVSKREGINAEQIVIKITDOGVY 59

Db 21 SYELGRYQAGQDKESNRVAYIDDGQAGKAENLTDPVSKREGINAEQIVIKITDOGVY 80

QY 60 TSHGDGHVHYNGKVPDYDAIISELLMKDPNYKLKDEDIVNEVKGKGVYKVDGKYVYLKD 119

Db 81 TSHGDGHVHYNGKVPDYDAIISELLMKDPNYQLKSDIVNEIKGGVIKVGKYYVYLKD 140

QY 120 AAHADNVRTKEEINROKQESHOREGTPNDGAVALARSQRYTTDDGYIFNASDIIE 179

Db 141 AAHADNIRTKEEIKRQERSHNHS---RADNAVAAARAGQRYTTDDGYIFNASDIIE 197

QY 180 TGDAYIVPHGDHVIYPKNELSAELAAEFALSGRNLSNRTYRRQNSDTSRTNWVP 239

Db 198 TGDAYIVPHGDHVIYPKNELSAELAAEFALSGRNLSNRTYRRQNSDTSRTNWVP 246

QY 240 SVSNPGTNTNTNNSNTNSQASQNDIDSLLKOLYKPLSORHVESDGLVFDPQAITSR 299

Db 247 NPAQRLSENHNLTVTTYHQ-NQGENISLLRELAKPLSERHVESDGLIFDPAQITSR 305

QY 300 TARGVAVPHGDHVIYPYSQMSLEERIAIIPLYRSNHVWPDSRPSPQPTPEPSP 359

Db 306 TARGVAVPHGNHYHFIPYEQMSLEERIAIIPLYRSNHVWPDSRPSPQPTPEPSP 365

QY 360 GPQAPNLKIDSNSLSVLVRKGVGYFEEKGISRYVFAKDLPSFTVNLESKLSKOE 419

Db 366 SPQPAPSNPID--GKLVKEAVRKVDGYFEEENGVSRYIIPAKOLSAFTAAGDISKLAKOE 423

QY 420 SVSHTLTAKENVAPROQEYFDKAYNLLTEAHKALFXNKGRNSDFOALDKLLBLRNDST 479

Db 424 SLSHKLGTKTKDTLSSPREFFYNKAYDULARIHQDLLNKGQVDFEALDNLLERLXOVES 483

QY 480 NKEVLVDLLAFAPITHPERLGPXNSQIEYTEDEVRIAGLADKYTTSDGYIFEDEHDITS 539

; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 388
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 194
 ; LENGTH: 826
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-09-769-787-194

Query Match 67.1%; Score 2795; DB 10; Length 826;
 Best Local Similarity 67.0%; Pred. No. 1.1e-192;
 Matches 549; Conservative 83; Mismatches 139; Indels 48; Gaps 8;
 QY 1 SYELGLYQA-RTVKENNRVSYIDKQATQKTENLTPDEVSKREGINAEOIVIKITDQGYV 59
 DB 21 SYELGRHQAGQDKKESNRVAYIDGQAGQKAENLTPDEVSKREGINAEOIVIKITDQGYV 80
 QY 60 TSHGDHYHYNGKVPYDAIISEELLMKDPNPKLDEIVNEVKGGYVVKDGYVYVLKD 119
 DB 81 TSHGDHYHYNGKVPYDAIISEELLMKDPNPKLDEIVNEVKGGYVVKDGYVYVLKD 140
 QY 120 AAHADNVRTKEEINRQKQHSOHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIID 179
 DB 141 AAHADNVRTKEEIKRQKQSHNH--GSGANDHAAVAAARAGRYTTDDGYIFNASDIID 198
 QY 180 TGDAYIVPHGDHYHYIPKNELSAELAAAFISGRGNLSNRYRQNSDNTSRTNWVP 239
 DB 199 TGDAYIVPHGDHYHYIPKNELSAELAAAFISGRGNLSNRYRQNSDNTSRTNWVP 239
 QY 240 SVSNPGTNTNTSNNSNTSQASQSDNDISLLKQLYKPLSORHVESDGLVDFDPAQITSR 299
 DB 248 NPAQPLSENHNLTVTPTYHQ--NOGENISLLRELAKPLSERHVESDGLVDFDPAQITSR 306
 QY 300 TARGVAVPHGDHYHYIPYQMSLEERARIIPLYRSNHWVDPDSRPEQSPQPTPEPSP 359
 DB 307 TARGVAVPHGNHYHYFIPYEQMSLEKRIARIIPLYRSNHWVDPDSRPEQSPQPTPEPSP 366
 QY 360 GPQAPNLK-IDSN---SSLVSQLVRKGGYVPEEKIGISRYPAKDLSETVKNLESKL 415
 DB 367 SPQAPNPQAPSNPIDKLVKAVRKVGQVFEENGVSRYIPAKDLSAETAAGIDSKL 426
 QY 416 SKQESVSHIT-TAKENAVAPDQDFYDKAYNLLTEAHKALFXNKGNSDFOALDKLLERLN 475
 DB 427 AKGESLHKLGAKTDLPSDRFYNKAYDLARIHQDLDNKGQVDFEALDNLLERLK 486
 QY 476 DESTNKEKLVDDLAFAPITHTPERLGKPNQSOIETDEVRIAQLADKYTTSDGYIFDEH 535
 DB 487 DVPSDKVLVDLAFAPIRHPRERLGKPNQSOIETDEVRIAQLADKYTTSDGYIFDPR 546
 QY 536 DIISDGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTKKGLPSPDADVKNPTGDS 595
 DB 547 DITSDEGDAYVTPHMTSHWIKDLSAEARAAQAYAKEKGLTPPSTDHQDSGNTKAG 606
 QY 596 AAAYNRVKEKRIPLVRPYMVEHTVEKNGNLIIPHDKDHYNIKPAFDDHYKAPNG 655
 DB 607 AEAYNRVAAKVPDRMPYNIQYTVKNGSLIIPHYDHYNIKFEWDEGLYEAPKG 666
 QY 656 YTELDLFAIKYVVEHPDRPHSDNCGWNASEHVLGKDHSDPNKNFKADEB-----P 709
 DB 667 YTELDLFAIKYVVEHPDRPHSDNCGWNASEHVLGKDHSDPNKNFKADEB-----P 726
 QY 710 VEET-----PAEVPQVETKEVQAQLKEAEVLLAKVTTSSSL 746
 DB 727 EEETPREEKQSEKPEPKPTPEPESPESEBPQVETKEVBEKLEAEADLLKIQDPII 786
 QY 747 KANATETLAGLNLLTLQIMDNNSIMAEKLLALLKGS 785
 DB 787 KSNKETLTGLKNLLFGTDQNNTIMAEKLLALLKGS 825

RESULT 10

US-10-412-862-10

; Sequence 10, Application US/10412862

; Publication No. US20040052781A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-685
 ; CURRENT APPLICATION NUMBER: US/10/412,862
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 10
 ; LENGTH: 819
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-10-412-862-10

Query Match 67.0%; Score 2788.5; DB 12; Length 819;
 Best Local Similarity 66.7%; Pred. No. 3.2e-192;
 Matches 544; Conservative 85; Mismatches 139; Indels 47; Gaps 7;
 QY 1 SYELGLYQA-RTVKENNRVSYIDKQATQKTENLTPDEVSKREGINAEOIVIKITDQGYV 59
 DB 21 SYELGRHQAGQDKKESNRVAYIDGQAGQKAENLTPDEVSKREGINAEOIVIKITDQGYV 80
 QY 60 TSHGDHYHYNGKVPYDAIISEELLMKDPNPKLDEIVNEVKGGYVVKDGYVYVLKD 119
 DB 81 TSHGDHYHYNGKVPYDAIISEELLMKDPNPKLDEIVNEVKGGYVVKDGYVYVLKD 140
 QY 120 AAHADNVRTKEEINRQKQHSOHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIID 179
 DB 141 AAHADNVRTKEEIKRQKQSHNH---RADNAAVAAARAGRYTTDDGYIFNASDIID 197
 QY 180 TGDAYIVPHGDHYHYIPKNELSAELAAAFISGRGNLSNRYRQNSDNTSRTNWVP 239
 DB 198 TGDAYIVPHGDHYHYIPKNELSAELAAAFISGRGNLSNRYRQNSDNTSRTNWVP 246
 QY 240 SVSNPGTNTNTSNNSNTSQASQSDNDISLLKQLYKPLSORHVESDGLVDFDPAQITSR 299
 DB 247 NPAQPLSENHNLTVTPTYHQ--NOGENISLLRELAKPLSERHVESDGLVDFDPAQITSR 305
 QY 300 TARGVAVPHGDHYHYFIPYQMSLEERARIIPLYRSNHWVDPDSRPEQSPQPTPEPSP 359
 DB 306 TARGVAVPHGNHYHYFIPYEQMSLEKRIARIIPLYRSNHWVDPDSRPEQSPQPTPEPSP 365
 QY 360 GPQAPNLKIDSNLSLVSQLVRKGGYVPEEKIGISRYPAKDLPSRTVKNLESKQKE 419
 DB 366 SPQAPSNPID--GKLVKAVRKVGQVFEENGVSRYIPAKDLSAETAAGIDSKLQKE 423
 QY 420 SVSHTLTAKENAVAPDQDFYDKAYNLLTEAHKALFXNKGNSDFOALDKLLERLNDEST 479
 DB 424 SLSHKLGTXTDLPSDRFYNKAYDLARIHQDLDNKGQVDFEALDNLLERLKDVSS 483
 QY 480 NKEKLVDDLAFAPITHTPERLGKPNQSOIETDEVRIAQLADKYTTSDGYIFDEHDIIS 539
 DB 484 DKVKLVVEDLAFAPIRHPRERLGKPNQSOIETDEVRIAQLADKYTTSDGYIFDPRDITS 543
 QY 540 DSGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTKKGLPSPDADVKNPTGDSAAAI 599
 DB 544 DSGDAYVTPHMTSHWIKDLSAEARAAQAYAKEKGLTPPSTDHQDSGNTKAGAEAI 603
 QY 600 YNRVKEKRIPLVRPYMVEHTVEKNGNLIIPHDKDHYNIKPAFDDHYKAPNGVYTL 659
 DB 604 YNRVAAKVPDRMPYNIQYTVKNGSLIIPHYDHYNIKFEWDEGLYEAPKYTL 663
 QY 660 DLFAIKYVVEHPDRPHSDNCGWNASEHVLGKDHSDPNKNFKADEB-----PVEET 713
 DB 664 DLLATVKYVVEHPDRPHSDNCGWNASEHVLGKDHSDPNKNFKADEB-----PVEET 723

```

QY 121 AHADNVRTKEEINRQKQHSQHREGCTPRNDGAVALARSGQRYTDDGQYIFNASDIIEDT 180
DB 141 AHADNVRTKEEINRQKQHSQHREGCTPRNDGAVALARSGQRYTDDGQYIFNASDIIEDT 200
QY 181 GDAYIVPHGDHYYHYPKNELSASELAFAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
DB 201 GDAYIVPHGDHYYHYPKNELSASELAFAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 260
QY 241 VSNPGTNTNTNNSNTNSQASQSDNDISLLKQLYKLPLSQRHVESDGLVFPDPAQITST 300
DB 261 VSNPGTNTNTNNSNTNSQASQSDNDISLLKQLYKLPLSQRHVESDGLVFPDPAQITST 320
QY 301 ARGVAVPHGDHYYHYPKNELSASELAFAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 360
DB 321 ARGVAVPHGDHYYHYPKNELSASELAFAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 380
QY 361 POPAPNLK-IDSN----SSIVSQLVRKVGEGYVFEKGIISRYVFAKDLPSSETVKNLESKLS 416
DB 381 POPAPNPQAPSNPIDKLVKEAVRKVGEGYVFEENGVSRYIPAKNLSAETAAGIDSKLA 440
QY 417 KOESVSHTLTAKENVAAPRQDFYKAYNLLTEAHKALFYKNGRNSDFOALDKLLERLND 476
DB 441 KOESLSHKLGAKTDLPSDDREFYKAYDILLARIHQDILLDNKGRQVDFEALDNLLERLND 500
QY 477 ESTNKEKLVDDLLAFIAPITHPERLGKPNLSQIEYTEDEVRIAQLADKYTTSDGYIFDEHD 536
DB 501 VSSDKVKLVDDLLAFIAPIRHPERLGKPNQAITYTDDDEIQVAKLAGKYTTEDGYIFDPRD 560
QY 537 IISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTEKGIILPSPDADVKANPTGDSA 596
DB 561 IISDEGDAYVTPHMTSHWIKKDSLSAEARAAQAYAKEKGLTPPSTDHQSNGTEAKGA 620
QY 597 AAIYNRVKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKHQHYHNIKFAWFDHDTYKAPNGY 656
DB 621 EAIYNRVKAAKVPLDRMPYNLYQYTVVEVKNGLIIPHQHYHNIKFEWDFEGLYEAPKGY 680
QY 657 TLEDLFATIKYVVEHPDERPHSDNGMGNASEHVLGKKHSDSDPNKFKAD----- 706
DB 681 TLEDLLATVKYVVEHPDERPHSDNGMGNASDHVQRNKGADNTQTEKSEKPKQTEKPE 740
QY 707 -----EFPVETPAEPVQVETEKVQAQLKEAEVLLAKVTDSLSL 746
DB 741 EETPREKQSEKPEPKPTEPERPEESPESEEPQVETEKVEKLEAREADLLGKIQDPII 800
QY 747 KANAFETIAGLNNLTQIMDNNSIMAEAEKLLALLKGS 785
DB 801 KSNABETLTLGKNLLFGTQDNNNTIMAEAEKLLALLKES 839

```

RESULT 8

```

US-10-324-143-38
; Sequence 38, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

FEATURE:

```

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-38

```

```

Query Match 71.1%; Score 2961; DB 15; Length 690;
Best Local Similarity 80.7%; Pred. No. 8.7e-205;
Matches 556; Conservative 54; Mismatches 75; Indels 4; Gaps 2;

QY 1 SYELGLYQARTVKENNRVSYIDGKQATOKTENLTTPDEVSKREGINAEIVIKITDQGYVT 60
DB 2 AYELGLHQATVKENNRVSYIDGKQATOKTENLTTPDEVSKREGINAEIVIKITDQGYVT 61
QY 61 SHGDHYYHNGKVPYDAIISBELLMKDPNYKLKDEIDVNEIKGGYVIVKDGKYYVYLKDA 120
DB 62 SHGDHYYHNGKVPYDAIISBELLMKDPNYQLKSDIIVNEIKGGYVIVKNGKYYVYLKDA 121
QY 121 AHADNVRTKEEINRQKQHSQHREGCTPRNDGAVALARSGQRYTDDGQYIFNASDIIEDT 180
DB 122 AHADNVRTKEEINRQKQHSQHREGCTPRNDGAVALARSGQRYTDDGQYIFNASDIIEDT 181
QY 181 GDAYIVPHGDHYYHYPKNELSASELAFAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
DB 182 GDAYIVPHGDHYYHYPKNELSASELAFAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 241
QY 241 VSNPGTNTNTNNSNTNSQASQSDNDISLLKQLYKLPLSQRHVESDGLVFPDPAQITST 300
DB 242 VSNPGTNTNTNNSNTNSQASQSDNDISLLKQLYKLPLSQRHVESDGLVFPDPAQITST 301
QY 301 ARGVAVPHGDHYYHYPKNELSASELAFAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 360
DB 302 ARGVAVPHGDHYYHYPKNELSASELAFAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 361
QY 361 POPAPNLK-IDSN----SSIVSQLVRKVGEGYVFEKGIISRYVFAKDLPSSETVKNLESKLS 416
DB 362 POPAPNPQAPSNPIDKLVKEAVRKVGEGYVFEENGVSRYIPAKNLSAETAAGIDSKLA 421
QY 417 KOESVSHTLTAKENVAAPRQDFYKAYNLLTEAHKALFYKNGRNSDFOALDKLLERLND 476
DB 422 KOESLSHKLGAKTDLPSDDREFYKAYDILLARIHQDILLDNKGRQVDFEALDNLLERLND 481
QY 477 ESTNKEKLVDDLLAFIAPITHPERLGKPNLSQIEYTEDEVRIAQLADKYTTSDGYIFDEHD 536
DB 482 VSSDKVKLVDDLLAFIAPIRHPERLGKPNQAITYTDDDEIQVAKLAGKYTTEDGYIFDPRD 541
QY 537 IISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTEKGIILPSPDADVKANPTGDSA 596
DB 542 IISDEGDAYVTPHMTSHWIKKDSLSAEARAAQAYAKEKGLTPPSTDHQSNGTEAKGA 601
QY 597 AAIYNRVKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKHQHYHNIKFAWFDHDTYKAPNGY 656
DB 602 EAIYNRVKAAKVPLDRMPYNLYQYTVVEVKNGLIIPHQHYHNIKFEWDFEGLYEAPKGY 661
QY 657 TLEDLFATIKYVVEHPDERPHSDNGMGNASEHVLGKKHSDSDPNKFKAD----- 690
DB 662 TLEDLLATVKYVVEHPDERPHSDNGMGNASEHVLGKKHSDSDPNKFKAD----- 690

```

RESULT 9

```

US-09-769-787-194
; Sequence 194, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164

```



```

; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; PRIOR FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-8

Query Match 100.0%; Score 4163; DB 15; Length 819;
Best Local Similarity 99.9%; Pred. No. 1.9e-291;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVT 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 21 SYELGLYQARTVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVT 80
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 SHGDHYHYNGKVPYDAIISBELLMKDPNYKDKDEIVNEVKGYYVIVKDGKYYVYLKDA 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 81 SHGDHYHYNGKVPYDAIISBELLMKDPNYKDKDEIVNEVKGYYVIVKDGKYYVYLKDA 140
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 141 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 200
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 181 GDAYIVPHGDHYHYTPKNELASSELAAAEAFISGRGNLSNSTRYRQNSDNTSRTNWVPS 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 201 GDAYIVPHGDHYHYTPKNELASSELAAAEAFISGRGNLSNSTRYRQNSDNTSRTNWVPS 260
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 241 VSNPGTTNTNNSNTNSQASQSDNDISLLKQLYKLPQSQRHVESDGLVFPDPAQITSR 300
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 261 VSNPGTTNTNNSNTNSQASQSDNDISLLKQLYKLPQSQRHVESDGLVFPDPAQITSR 320
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 301 ARGVAVPHGDHYHYTFPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 360
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 321 ARGVAVPHGDHYHYTFPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 380
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 361 POPAPNLKIDSNSLSVQLVRKVGEGYVEEKGISRYVFAKOLPSETVKNLESKLSKQBS 420
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 381 POPAPNLKIDSNSLSVQLVRKVGEGYVEEKGISRYVFAKOLPSETVKNLESKLSKQBS 440
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 421 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 480
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 441 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 500
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 481 KEKLVDLLAFIAPITHPERLKQKPSQIBYTEDEVRIQAQADKYTTSDGYIFDEHDIISD 540
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 501 KEKLVDLLAFIAPITHPERLKQKPSQIBYTEDEVRIQAQADKYTTSDGYIFDEHDIISD 560
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 541 EGDAYVTHMGSHHWIGKDSLSKKEKVAQAQYTKKGIILPPSPDADVKNPPTGDSAAALY 600
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 561 EGDAYVTHMGSHHWIGKDSLSKKEKVAQAQYTKKGIILPPSPDADVKNPPTGDSAAALY 620
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 601 NRVKGEKRIPLVRLPVMVEHTVEVKNGLIIPKHQHYHNIKEAWPDHHTYKAPNGVTLED 660
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 621 NRVKGEKRIPLVRLPVMVEHTVEVKNGLIIPKHQHYHNIKEAWPDHHTYKAPNGVTLED 680
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 661 LPATIKYVVEHDPDRPHSDNGMGNASEHVLGKKDHSDEPNKNFKADEEPEETPAEPEVP 720
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 681 LPATIKYVVEHDPDRPHSDNGMGNASEHVLGKKDHSDEPNKNFKADEEPEETPAEPEVP 740
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 721 QVETEKVEAQLKEAEVILLAKVTDSLSKANATETLAGLRNLLTQIMDNNSIMAEAKLLA 780
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 741 QVETEKVEAQLKEAEVILLAKVTDSLSKANATETLAGLRNLLTQIMDNNSIMAEAKLLA 800
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```

QY 781 LLKGSNPSSVSKKIN 796
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 801 LLKGSNPSSVSKKIN 816
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 5
US-10-324-143-19
; Sequence 19, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-19

Query Match 77.3%; Score 3218; DB 15; Length 821;
Best Local Similarity 75.2%; Pred. No. 3.2e-223;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVT 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 2 AVELGLHQARTVKENNRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVT 61
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 SHGDHYHYNGKVPYDAIISBELLMKDPNYKDKDEIVNEVKGYYVIVKDGKYYVYLKDA 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 62 SHGDHYHYNGKVPYDAIISBELLMKDPNYQLKDSIDIINEIKGGYIVKNGKYYVYLKDA 121
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 122 AHADNVRTKEEINRQKQHSQHREGGTSANDGAVAFARSQGRYTTDDGYIFNASDIIEDT 181
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 181 GDAYIVPHGDHYHYTPKNELASSELAAAEAFISGRGNLSNSTRYRQNSDNTSRTNWVPS 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 182 GDAYIVPHGDHYHYTPKNELASSELAAAEAFISGRGNLSNSTRYRQNSDNTSRTNWVPS 241
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 241 VSNPGTTNTNNSNTNSQASQSDNDISLLKQLYKLPQSQRHVESDGLVFPDPAQITSR 300
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 242 VSNPGTTNTNNSNTNSQASQSDNDISLLKQLYKLPQSQRHVESDGLVFPDPAQITSR 301
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 301 ARGVAVPHGDHYHYTFPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 360
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 302 ARGVAVPHGNHYHYTFPEQMSLEKRIARIIPLYRSNHWVPSRPEPSPQPTPEPSPS 361
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 361 POPAPNLK-IDSN---SSIVSOLVRKVGEGYVEEKGISRYVFAKOLPSETVKNLESKLS 416
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 362 POPAPNPQAPSNPIDKLVKEAVRKGVDGVEENGVSRYIPAKNLSAETAAGIDSKLA 421
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 417 KOESVSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLN 476
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 422 KOESLSHLKGAKKTDLPSSDREFYNKAYDILLARIHQDLDLNDKGRQVDFEALDNLRLK 481
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 477 ESTNKEKLVDDLLAPAPIIPIRPERLGKNSQIETEVDEVRITQAQADKYTTSDGYIFD 536
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 482 VSSDKVKLVDDILLAPAPIRPIRPERLGKPNQAQITTTDDDEIQVAKLAGKYTTEDGYIF 541
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```

Db      81  SHGDHYHYNGKVPYDAIISSELLMKDPNYLKDEDIVNEVKGGVVVKDGGYVYVLKDA 140
      121  AHADNVTKEINRQKQHSQHRGGTTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
Db      141  AHADNVTKEINRQKQHSQHRGGTTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 200
      181  GDYIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Db      201  GDYIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 260
      241  VSNPGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 300
Db      261  VSNPGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 320
      301  ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
Db      321  ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 380
      361  POPAPNLKIDSNLSVLQVRKVGEGYVFEKGIISRYVFAKDLPSSETVKNLESKLSQES 420
Db      381  POPAPNLKIDSNLSVLQVRKVGEGYVFEKGIISRYVFAKDLPSSETVKNLESKLSQES 440
      421  VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db      441  VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
      481  KEKLVDDLLAFIAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEHDIISD 540
Db      501  KEKLVDDLLAFIAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEHDIISD 560
      541  EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPPSPADVKANPTGDSAAAIY 600
Db      561  EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPPSPADVKANPTGDSAAAIY 620
      601  NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFADFDDHTYKAPNGYTTLED 660
Db      621  NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFADFDDHTYKAPNGYTTLED 680
      661  LFATIKYVVEHPDERPHSDNGWGNASBHVLGKDHSDPNKFNKADBEPEETPAEPVP 720
Db      681  LFATIKYVVEHPDERPHSDNGWGNASBHVLGKDHSDPNKFNKADBEPEETPAEPVP 740
      721  QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 780
Db      741  QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 800
      781  LLKGSNPFSSVSKEKIN 796
Db      801  LLKGSNPFSSVSKEKIN 816

```

```

RESULT 3
US-10-412-850-8
; Sequence 8, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 8
; LENGTH: 819

```

```

; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-8

Query Match      100.0%; Score 4163; DB 15; Length 819;
Best Local Similarity 99.9%; Pred. No. 1.9e-291;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  SYELGLCAARTVKNENRVSVIDGQATQCTENLTNPDEVSKREGINAEQIVIKITDQGVVT 60
Db      21  SYELGLCAARTVKNENRVSVIDGQATQCTENLTNPDEVSKREGINAEQIVIKITDQGVVT 80
      61  SHGDHYHYNGKVPYDAIISSELLMKDPNYLKDEDIVNEVKGGVVVKDGGYVYVLKDA 120
Db      81  SHGDHYHYNGKVPYDAIISSELLMKDPNYLKDEDIVNEVKGGVVVKDGGYVYVLKDA 140
      121  AHADNVTKEINRQKQHSQHRGGTTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
Db      141  AHADNVTKEINRQKQHSQHRGGTTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 200
      181  GDYIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Db      201  GDYIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 260
      241  VSNPGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 300
Db      261  VSNPGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 320
      301  ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
Db      321  ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 380
      361  POPAPNLKIDSNLSVLQVRKVGEGYVFEKGIISRYVFAKDLPSSETVKNLESKLSQES 420
Db      381  POPAPNLKIDSNLSVLQVRKVGEGYVFEKGIISRYVFAKDLPSSETVKNLESKLSQES 440
      421  VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db      441  VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
      481  KEKLVDDLLAFIAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEHDIISD 540
Db      501  KEKLVDDLLAFIAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEHDIISD 560
      541  EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPPSPADVKANPTGDSAAAIY 600
Db      561  EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPPSPADVKANPTGDSAAAIY 620
      601  NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFADFDDHTYKAPNGYTTLED 660
Db      621  NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFADFDDHTYKAPNGYTTLED 680
      661  LFATIKYVVEHPDERPHSDNGWGNASBHVLGKDHSDPNKFNKADBEPEETPAEPVP 720
Db      681  LFATIKYVVEHPDERPHSDNGWGNASBHVLGKDHSDPNKFNKADBEPEETPAEPVP 740
      721  QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 780
Db      741  QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 800
      781  LLKGSNPFSSVSKEKIN 796
Db      801  LLKGSNPFSSVSKEKIN 816

RESULT 4
US-10-387-783-8
; Sequence 8, Application US/10387783
; Publication No. US2004000531A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

```

89 649 15.6 272 15 US-10-324-143-2 Sequence 2, Appli
90 649 15.6 908 15 US-10-324-143-114 Sequence 114, App

ALIGNMENTS

```
RESULT 1
US-09-765-272-56
; Sequence 56, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-765-272-56

Query Match 100.0%; Score 4163; DB 9; Length 796;
Best Local Similarity 100.0%; Pred. No. 1.8e-291;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
DB 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGGYVIVKDGKYYVYLKDA 120
DB 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGGYVIVKDGKYYVYLKDA 120
QY 121 AHADNVRTKEENRQKQESHQREGTTPNDGAVALARSQGRYTTDDGYIFNASDIIEDT 180
DB 121 AHADNVRTKEENRQKQESHQREGTTPNDGAVALARSQGRYTTDDGYIFNASDIIEDT 180
QY 181 GDAYIVPHGDHYHYIPKXNELSASELAAAFUSGRGNLSNSTRYRQNSDNTSRTNWVPS 240
DB 181 GDAYIVPHGDHYHYIPKXNELSASELAAAFUSGRGNLSNSTRYRQNSDNTSRTNWVPS 240
QY 241 VSNPGTNTNTNSNTNSQASQSDNIDSLKQLYKPLSQRHVESDGLVDFDPAQITSR 300
DB 241 VSNPGTNTNTNSNTNSQASQSDNIDSLKQLYKPLSQRHVESDGLVDFDPAQITSR 300
```

```
DB 241 VSNPGTNTNTNSNTNSQASQSDNIDSLKQLYKPLSQRHVESDGLVDFDPAQITSR 300
QY 301 ARGVAVPHGDHYHYIPYSQMSLEERIARIIPLYRNSHNWVPSRPEQSPQPTPEPSPG 360
DB 301 ARGVAVPHGDHYHYIPYSQMSLEERIARIIPLYRNSHNWVPSRPEQSPQPTPEPSPG 360
QY 361 POPAPNLKIDNSLSLVQVRKVGEGYVFEKSGISRYVFAKDLPSVTKNLESKLSQES 420
DB 361 POPAPNLKIDNSLSLVQVRKVGEGYVFEKSGISRYVFAKDLPSVTKNLESKLSQES 420
QY 421 VSHTLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
DB 421 VSHTLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
QY 481 KEKLVDDLAFAPITPHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDDEHDIISD 540
DB 481 KEKLVDDLAFAPITPHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDDEHDIISD 540
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKEGILPPSPDADVKANPTGDSAAAIY 600
DB 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKEGILPPSPDADVKANPTGDSAAAIY 600
QY 601 NRVKGEKRIPLVRLPYMVVEHTVEVKNGLIIPHKDHYHNIKFAWFDDHTYKAPNGYTTLED 660
DB 601 NRVKGEKRIPLVRLPYMVVEHTVEVKNGLIIPHKDHYHNIKFAWFDDHTYKAPNGYTTLED 660
QY 661 LFAIKYVVEHPDPRPHSDNGWGNASEHVLGKDHSDPNKFNKADPEPVEETPAEPEVP 720
DB 661 LFAIKYVVEHPDPRPHSDNGWGNASEHVLGKDHSDPNKFNKADPEPVEETPAEPEVP 720
QY 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780
DB 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780
QY 781 LLKGSNPSSVSKEKIN 796
DB 781 LLKGSNPSSVSKEKIN 796

RESULT 2
US-10-412-862-8
; Sequence 8, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-8

Query Match 100.0%; Score 4163; DB 12; Length 819;
Best Local Similarity 99.9%; Pred. No. 1.9e-291;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
DB 21 SYELGLYQARTVKENNRVSYIDGKQATKTENLTPEVSKREGINAEQIVIKITDQGYVT 80
QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGGYVIVKDGKYYVYLKDA 120
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:16:48 ; Search time 405 Seconds

(without alignments)
632.474 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYELGLYQARTVKNRVS.....KLALLKGNPSSVSKKIN 756

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/prodata/2/pubppa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubppa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubppa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubppa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubppa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubppa/PCRS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubppa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubppa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubppa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubppa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubppa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubppa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubppa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubppa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubppa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubppa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubppa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 4163 | 100.0 | 796 | 9 | US-09-765-272-56 |
| 2 | 4163 | 100.0 | 819 | 12 | US-10-412-862-8 |
| 3 | 4163 | 100.0 | 819 | 15 | US-10-412-850-8 |
| 4 | 4163 | 100.0 | 819 | 15 | US-10-387-783-8 |
| 5 | 3218 | 77.3 | 821 | 15 | US-10-324-143-19 |
| 6 | 3218 | 77.3 | 840 | 10 | US-09-884-465A-7 |
| 7 | 3218 | 77.3 | 840 | 10 | US-10-324-143-8 |
| 8 | 2961 | 71.1 | 690 | 15 | US-10-324-143-38 |
| 9 | 2795 | 67.1 | 826 | 12 | US-09-769-787-194 |
| 10 | 2788.5 | 67.0 | 819 | 12 | US-10-412-862-10 |
| 11 | 2788.5 | 67.0 | 819 | 15 | US-10-412-850-10 |
| 12 | 2788.5 | 67.0 | 819 | 15 | US-10-387-783-10 |
| 13 | 2772 | 66.6 | 838 | 12 | US-10-412-862-4 |
| 14 | 2772 | 66.6 | 838 | 15 | US-10-412-850-4 |
| 15 | 2772 | 66.6 | 838 | 15 | US-10-387-783-4 |

| | | | | | | |
|----|--------|------|------|----|--------------------|--------------------|
| 16 | 2750.5 | 66.0 | 819 | 15 | US-10-324-143-32 | Sequence 32, Appl |
| 17 | 2750.5 | 66.0 | 838 | 10 | US-09-884-465A-8 | Sequence 8, Appl |
| 18 | 2750.5 | 66.0 | 838 | 15 | US-10-324-143-9 | Sequence 9, Appl |
| 19 | 2649.5 | 63.6 | 763 | 9 | US-09-765-272-66 | Sequence 66, Appl |
| 20 | 2170 | 52.1 | 613 | 15 | US-10-324-143-22 | Sequence 22, Appl |
| 21 | 1899 | 45.6 | 1126 | 10 | US-09-884-465A-383 | Sequence 383, App |
| 22 | 1898 | 45.6 | 1126 | 15 | US-10-324-143-54 | Sequence 54, Appl |
| 23 | 1891.5 | 45.4 | 1365 | 10 | US-09-884-465A-382 | Sequence 382, App |
| 24 | 1891.5 | 45.4 | 1365 | 15 | US-10-324-143-53 | Sequence 53, Appl |
| 25 | 1888 | 45.3 | 555 | 15 | US-10-324-143-26 | Sequence 26, Appl |
| 26 | 1808 | 43.4 | 612 | 15 | US-10-324-143-44 | Sequence 44, Appl |
| 27 | 1795.5 | 43.1 | 568 | 15 | US-10-324-143-33 | Sequence 33, App |
| 28 | 1795.5 | 43.1 | 1139 | 10 | US-09-884-465A-380 | Sequence 380, App |
| 29 | 1795.5 | 43.1 | 1139 | 15 | US-10-324-143-51 | Sequence 51, Appl |
| 30 | 1795.5 | 43.1 | 1378 | 10 | US-09-884-465A-378 | Sequence 378, App |
| 31 | 1795.5 | 43.1 | 1378 | 15 | US-10-324-143-49 | Sequence 49, Appl |
| 32 | 1666 | 40.0 | 334 | 15 | US-10-324-143-20 | Sequence 20, Appl |
| 33 | 1665 | 40.0 | 1238 | 10 | US-09-884-465A-381 | Sequence 381, App |
| 34 | 1661 | 39.9 | 1238 | 15 | US-10-324-143-52 | Sequence 52, Appl |
| 35 | 1656 | 39.8 | 999 | 10 | US-09-884-465A-376 | Sequence 376, App |
| 36 | 1655 | 39.7 | 999 | 15 | US-10-324-143-47 | Sequence 47, Appl |
| 37 | 1651 | 39.6 | 999 | 15 | US-10-324-143-48 | Sequence 48, Appl |
| 38 | 1649 | 39.6 | 428 | 15 | US-10-324-143-27 | Sequence 27, Appl |
| 39 | 1649 | 39.6 | 999 | 10 | US-09-884-465A-377 | Sequence 377, App |
| 40 | 1614.5 | 38.8 | 473 | 15 | US-10-324-143-36 | Sequence 36, Appl |
| 41 | 1591.5 | 38.2 | 1058 | 15 | US-10-324-143-46 | Sequence 46, Appl |
| 42 | 1552 | 37.3 | 487 | 15 | US-10-324-143-21 | Sequence 21, Appl |
| 43 | 1552 | 37.3 | 487 | 15 | US-10-324-143-35 | Sequence 35, Appl |
| 44 | 1247 | 29.9 | 780 | 15 | US-10-324-143-37 | Sequence 37, Appl |
| 45 | 1247 | 29.9 | 1019 | 15 | US-10-324-143-14 | Sequence 14, Appl |
| 46 | 1247 | 29.9 | 1039 | 10 | US-09-884-465A-6 | Sequence 6, Appl |
| 47 | 1247 | 29.9 | 1039 | 15 | US-10-324-143-7 | Sequence 7, Appl |
| 48 | 1217.5 | 29.2 | 489 | 15 | US-10-324-143-15 | Sequence 15, Appl |
| 49 | 1217.5 | 29.2 | 509 | 15 | US-10-324-143-16 | Sequence 16, Appl |
| 50 | 1203.5 | 28.9 | 484 | 10 | US-09-769-787-38 | Sequence 38, Appl |
| 51 | 1203.5 | 28.9 | 484 | 12 | US-10-412-862-6 | Sequence 6, Appl |
| 52 | 1203.5 | 28.9 | 484 | 15 | US-09-769-787-18 | Sequence 18, Appl |
| 53 | 1203.5 | 28.9 | 484 | 15 | US-10-387-783-6 | Sequence 6, Appl |
| 54 | 1187.5 | 28.5 | 447 | 9 | US-09-765-272-182 | Sequence 182, App |
| 55 | 991.5 | 23.8 | 342 | 15 | US-10-324-143-40 | Sequence 40, Appl |
| 56 | 991.5 | 23.8 | 913 | 10 | US-09-884-465A-384 | Sequence 384, App |
| 57 | 991.5 | 23.8 | 913 | 15 | US-10-324-143-55 | Sequence 55, Appl |
| 58 | 991.5 | 23.8 | 913 | 15 | US-09-884-465A-379 | Sequence 379, App |
| 59 | 991.5 | 23.8 | 1152 | 10 | US-10-324-143-50 | Sequence 50, Appl |
| 60 | 966.5 | 23.2 | 840 | 10 | US-09-884-465A-10 | Sequence 10, Appl |
| 61 | 966.5 | 23.2 | 840 | 10 | US-09-769-736-18 | Sequence 18, Appl |
| 62 | 929 | 22.3 | 822 | 10 | US-09-252-088-15 | Sequence 15, Appl |
| 63 | 927 | 22.3 | 793 | 10 | US-10-340-792-15 | Sequence 15, Appl |
| 64 | 927 | 22.3 | 793 | 15 | US-10-324-143-39 | Sequence 39, Appl |
| 65 | 925.5 | 22.2 | 294 | 15 | US-10-324-143-18 | Sequence 18, Appl |
| 66 | 865 | 20.8 | 205 | 15 | US-10-324-143-30 | Sequence 30, Appl |
| 67 | 853 | 20.5 | 226 | 15 | US-09-769-736-24 | Sequence 24, Appl |
| 68 | 841 | 20.2 | 381 | 10 | US-10-324-143-31 | Sequence 31, Appl |
| 69 | 802 | 19.3 | 203 | 15 | US-10-324-143-112 | Sequence 112, App |
| 70 | 738 | 17.7 | 908 | 15 | US-10-324-143-113 | Sequence 113, App |
| 71 | 738 | 17.7 | 908 | 15 | US-10-324-143-110 | Sequence 110, App |
| 72 | 733 | 17.6 | 908 | 15 | US-10-324-143-111 | Sequence 111, App |
| 73 | 733 | 17.6 | 906 | 10 | US-09-884-465A-369 | Sequence 369, App |
| 74 | 723.5 | 17.4 | 906 | 10 | US-09-884-465A-373 | Sequence 373, App |
| 75 | 714.5 | 17.2 | 906 | 10 | US-09-884-465A-371 | Sequence 371, App |
| 76 | 710.5 | 17.1 | 906 | 10 | US-10-324-143-157 | Sequence 157, App |
| 77 | 685.5 | 16.5 | 938 | 15 | US-10-324-143-90 | Sequence 90, Appl |
| 78 | 684 | 16.4 | 902 | 15 | US-10-324-143-151 | Sequence 151, App |
| 79 | 680.5 | 16.3 | 929 | 15 | US-10-324-143-153 | Sequence 153, App |
| 80 | 675.5 | 16.2 | 900 | 10 | US-09-884-465A-333 | Sequence 333, App |
| 81 | 665.5 | 16.0 | 925 | 15 | US-10-324-143-152 | Sequence 152, App |
| 82 | 660 | 15.8 | 715 | 10 | US-09-252-088-16 | Sequence 16, Appl |
| 83 | 660 | 15.8 | 715 | 15 | US-10-340-792-16 | Sequence 16, Appl |
| 84 | 659.5 | 15.8 | 929 | 15 | US-10-324-143-150 | Sequence 150, App |
| 85 | 655 | 15.7 | 917 | 15 | US-10-324-143-156 | Sequence 156, App |
| 86 | 654.5 | 15.7 | 907 | 15 | US-10-324-143-158 | Sequence 158, App |
| 87 | 650 | 15.6 | 906 | 15 | US-10-324-143-3 | Sequence 3, Appl |
| 88 | 650 | 15.6 | 908 | 15 | US-10-324-143-149 | Sequence 149, Appl |


```

; Patent No. 6455026
; GENERAL INFORMATION:
; APPLICANT: Melcher, Thorsten
; APPLICANT: Mueller, Sabine
; APPLICANT: Chin, Daniel
; TITLE OF INVENTION: Use of Protein Tyrosine Phosphatase Zeta as a Biomolecular Target
; TITLE OF INVENTION: Treatment and Visualization of Brain Tumors
; FILE REFERENCE: 262/235 ACY
; CURRENT APPLICATION NUMBER: US/09/816, 703A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Krueger, N.X. and Saito, H.: A human transmembrane protein-tyrosi
; OTHER INFORMATION: ne-phosphatase, PTP zeta, is expressed in brain and has an N-term
; OTHER INFORMATION: inal receptor domain homologous to carbonic anhydrases; Proc. Nat
; OTHER INFORMATION: l. Acad. Sci. USA 89 (16), 7417-7421 (1992)
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Levy, J.B., et al.: The cloning of a receptor-type protein tyrosi
; OTHER INFORMATION: ne phosphatase expressed in the central nervous system; J. Biol.
; OTHER INFORMATION: Chem. 268 (14), 10573-10581 (1993)
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION: By similarity
; NAME/KEY: DOMAIN
; LOCATION: (25)..(1635)
; OTHER INFORMATION: Extracellular (potential)
; NAME/KEY: misc.feature
; LOCATION: (25)..(2314)
; OTHER INFORMATION: Mature chain; protein-tyrosine phosphatase zeta
; NAME/KEY: DOMAIN
; LOCATION: (34)..(302)
; OTHER INFORMATION: Carbonic-anhydrase like
; NAME/KEY: SITE
; LOCATION: (105)..(105)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (134)..(134)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (223)..(223)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (232)..(232)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: DOMAIN
; LOCATION: (312)..(406)
; OTHER INFORMATION: Fibronectin Type-III
; NAME/KEY: SITE
; LOCATION: (324)..(324)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (381)..(381)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (497)..(497)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (501)..(501)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (552)..(552)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (587)..(587)
; OTHER INFORMATION: Chondroitin sulfate (potential)

```

```

; NAME/KEY: SITE
; LOCATION: (602)..(602)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (629)..(629)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (637)..(637)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (677)..(677)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: VARSPLIC
; LOCATION: (755)..(1614)
; OTHER INFORMATION: Splicing variant; missing (in short isoform)
; NAME/KEY: BINDING
; LOCATION: (997)..(997)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1017)..(1017)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1050)..(1050)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1082)..(1082)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1122)..(1122)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1456)..(1456)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (1548)..(1548)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: BINDING
; LOCATION: (1550)..(1550)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1561)..(1561)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC ...) (potential)
; NAME/KEY: SITE
; LOCATION: (1617)..(1617)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC ...) (potential)
; NAME/KEY: TRANSMEM
; LOCATION: (1636)..(1661)
; OTHER INFORMATION: Transmembrane region; potential
; NAME/KEY: DOMAIN
; LOCATION: (1662)..(2314)
; OTHER INFORMATION: Cytoplasmic (potential)
; NAME/KEY: CONFLICT
; LOCATION: (1722)..(1728)
; OTHER INFORMATION: Missing (in ref. 2)
; NAME/KEY: DOMAIN
; LOCATION: (1744)..(1997)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: ACT_SITE
; LOCATION: (1932)..(1932)
; OTHER INFORMATION: Active site; by similarity
; NAME/KEY: DOMAIN
; LOCATION: (1998)..(2314)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: misc.feature
; LOCATION: (2222)..(2222)
; OTHER INFORMATION: Ancestral active site
; US-09-816-703A-2

```

Query Match 3.4%; Score 141.5; DB 4; Length 2314;

Best Local Similarity 19.2%; Pred. No. 0.14; Gaps 43;

Matches 169; Conservative 113; Mismatches 267; Indels 331; Gaps 43;

QY 51 IKITDQGYVTSHGDHYHYNGKVPYDAIISFELLMKDPNKLKDEIVNEVKGSGYIKVD 110

Db 212 PEYKLETKGTQEPGHEGAAVREDLPVYTKPLETKGTQPGHEGAAVREBPAYTER 271
Qy 299 RTARGVAVGHGDIHYFIPYSOMSELEERARIIPLYRSNHWVDSRPEQSPQTPPEPS 358
Db 272 LATKGQEPGHE-----GKATVREETLEYTEFPVATKGT-----QEPHEGEXRVEEEL 319
Qy 359 PGOPPA-----PML-----KIDNSSLSVLVYR--KVGEVYFEE 391
Db 320 PALEVTRNRTETONIPYETEEIQDPTLLKXNRKIEROQAGRTTIQYEDIYVNGNVET 379
Qy 392 KGHSRYVFAKDLSETVKNLESKLSQESVSHLTAKENAVAPRDOEFYDKAYNLL--TE 449
Db 380 KEVSRTEVA--PVNEVVKVGLVVKPTEIINLTKVEN-----KKSITVSYNLIDTTS 431
Qy 450 AHKALFXNKRNSDFQALDKLLERLNDESTNKEKLVDDLLAFAPL--TH--PERLGKPN 505
Db 432 AYVS-----AKTVFHHG-DKLKVEVDIENPAKEQVTSGLDYVTPYVTKHLTYNLGNN 484
Qy 506 SQEYTEDVRIQAQADKYTSDGYTFDEHDIISDEGDAYVTPHMGHSHWIGKDSLSKXE 565
Db 485 E-ENTETSDQL--EYKIEIKDIDSVELYKENDRY-----RRYL--SLEAP 530
Qy 566 KVAQAAYTKKLGILPPSPADVKANPTGDSAAAAYNRVKGKRIPLVRLPYVVEHT 621
Db 531 TDTAKYFVKV-----SDRFKEMVLPVKS-----ITENTDGT 563
Qy 622 -VEVKNGLIIPKDDHYNIKFAWFDHY-----KAPNGYT-LEDLPATIK-----Y 667
Db 564 KVTVAVDQLVEGTGDYK-----DDYFTVAKSAEQGVYTSFKQLVTAMQNSLSGV 616
Qy 668 YVEHPDRPHSNDGWNAGSEHVLGKKDHS-----EDPNKNFKADE--EPVEET 713
Db 617 YTLASDM-----TAEVSLGDKQTSYLTGAFGLSGDGTSGYAIYDLKKELFD 667
Qy 714 PABPEVPQVETEKVEAQLKEAEVLLAKVTDSSL-----KANATETLAGLRNNLTQI 765
Db 668 LMGATVRDLDTKVSADSKENVAALAKAANSANNVAVBGKISGAKSVAGLASATNTV 727
Qy 766 MNNNSIMAEKLLALLKGSN 786
Db 728 IENSTFG---KLIAHQDSN 745

RESULT 22

US-09-200-650E-5
; Sequence 5, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Eikhinn, Deirdre N.
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1315
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-200-650E-5

Query Match 3.4%; Score 141.5; DB 4; Length 1315;
Best Local Similarity 18.5%; Pred. No. 0.057;
Matches 179; Conservative 128; Mismatches 327; Indels 331; Gaps 43;

Qy 12 VKENNRVSI-----DGKQATOKTENLTDPDEVSKREGINAEOIIVIKITDOGYVTSBGHDY 66
Db 160 LQENKSVVAVQPTNEENKKVDAKTESTT-----LVKSDALKSNDETLVDNNSN 210
Qy 67 HYYN-----GKVPYDAIISEELM-----KDPNYK--- 91
Db 211 NENNADIILPKSTAPKRLNTRMRIAIAVQPSSTSEAKNVNDLITSNTTLTVVDADKNKIVP 270
Qy 92 -----LKDEDIYNE-VKGG--YVIKVDGKYVYLKDAHADNVRKKEINRQKQESQH 142
Db 271 AQDYLKSKQITVDDKVKSGDYFTIKYSPTQVY--GLNPEDIKNIGDIK----- 318
Qy 143 REGGTNRNCAVALARSQGRYTTDDGVIFNASDIIIEDTGDAYI-----VPHGDHYH--- 194
Db 319 ----DNNGETIATAKHD-----TANNLIYTFDYVDRFNSVQMGINSIYMD 363
Qy 195 -----IPKNELASELAAAEAFLSGRGNLSNSTRQ-----NSDNTGRTNWWPSVS- 242
Db 364 ADTIPVSKNDVE-----FNVTIGNTTTKTANIQPDYVNVNEKNSIGSAFTETVSH 414
Qy 243 ----NPG-----TTNTNTSNNSTNSQ-----ASQSDIDSILK-----QYKLPESQ 281
Db 415 VGNKENPGYKQTIYVNPSENSLTNAKLKQVAHSSYPNNGINQINKDVTDIKIYQVP--K 472
Qy 282 RHVESDGLVDPQAQITSRARGV-AVPHGDHYHFIPYSOMSELEERARIIPLYRSNHW 340
Db 473 GYTLANKGYDVTNKELIDVTNQYLQKITYGDNNSAV--IDFGNADSAVVMVNTKFOYTN- 529
Qy 341 VPSRPEQSPQPTPEPSPGPAPNLKIDSNSL-----VSQLVKRVGEGYVFE- 390
Db 530 -----SESPTLVQMATLSGTGNKSVSTGNALGFTNNSQSGAGQEVYKIG-NYVWED 579
Qy 391 -----EKGISRYV-----FAKDLP-----SETVK 409
Db 580 TNKNGVQELGKGVGNVTVTVDNNTNTKVGEAVTKEDGSLYLPNLPNGDYRVEFSNLPK 639
Qy 410 NLESKLSKQ-----ESVSHLTAKENAVAPRDOEFYDKAYNLLTEAHKALFXNKGKN 461
Db 640 GYEVTPSKQGNBELDSNGLSSVITNGKDNLSADLGIYKPKYNLGDYVWEDTNKNGIQD 699
Qy 462 SDFQALDKLLERLNDESTNKEKLV---DLLAFAPLITHPERLKGPNKSOIYEYTED- 514
Db 700 QDEKGISGVTVTLKDENGVNLKTVTTDADGKYF-----TDLNGNYVVEETTPREGYT 752
Qy 515 -VRIAQLADKYTTSDGY-----IFEHDIISDEGDAYVTPHMGHSHWIGKDSLSKXVAA 569
Db 753 PTTVTSGSDIEKDSNGLTTTGTGVINGADNMWTLDSGFYKTPKYNLGNVYVWEDTNKDGQDST 812
Qy 570 QAYTKEKGLPPSPDADVKANPTGDSAAAAYNRVKGKRIPLVRLPYVWVHTVEVKNGNL 629
Db 813 ----EKGI-----SG-----VTVLKNENGVEV 830
Qy 630 IIPHK--DHYHNKPAWFDHYTK---APNGYTLDELFAFIKYVVEHPDRPHSNDGWN 684
Db 831 LQTKTKDKGKYQFTGLENGTYKVEPETSPTQVGS-----GTDEGIDSN---GT 880
Qy 685 ASEHVLGKKDHS-----EDPNKNFKADEEPVEETPAEPVQVETEK 726
Db 881 STTGVIKDKNDNTIDSGFYKPTNLDGYVWEDTNKNGVQDK----- 922
Qy 727 VEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTQIMDNNSIMAEKLLALLKGSN 786
Db 923 -EKGISGVTVTLKDENDKVLKVTITD-----ENKGYQFTDLNNGTYKVE--FETPSGYT 973
Qy 787 PSSVS 791
Db 974 PTVSV 978

RESULT 23
US-09-816-703A-2
; Sequence 2, Application US/09816703A

QY 497 -HPERLGKPN--OIEVTEDE-----VR--IAQL 520
Db 694 MSSQKLGNGISSSELIELQKMKESYRQVQELRSLYNLQTHESQKELMYGVNDIDAL 753
QY 521 ADKYTTS--DGYIFDEHDIISDEGDYVTPMHGSHWIGKDSLSDEKVAQAAYTKKGI 578
Db 754 VKTCTTSLNDADII--LSDYISDQSKFESQKQDIIANIGKIVSNFLQESLYTK-----808
QY 579 LPSPDADVANKPTGDSAAAI--YNRVKGKRIPLVR-----LPMVHEHTVEVK 625
Db 809 -----ADILHSHLNDTNSIRKANEIMNRSEEFLENAASQAEIVGANKERIOKTVE--860
QY 626 NGMLIIIPKDH-VHNIKFAFDDHTYKAPNG-----VTLEDLFATIKYVVEHPDERP 676
Db 861 NGSQKLSKSKAHSNSRSMYDCHLALAESQKQGVNLEVTDLRLLOKVK-----E 911
QY 677 HSDNGWGNASEHVLGKDHSEDPNKNFKADEEPVEETPABPEVQVETKEVQAQKBAEV 736
Db 912 HSEDNTEKEHQQLLDLLESVGNNDNL-----IDSIKTPHTELQ-----950
QY 737 LLAKVTDSLK-----ANATETLAGLN-----NLTLOIMDNNSIMAAE-----EKLLA 780
Db 951 ---KITDHLVKGTTSLANHTNELLGLGDESLCNLETTIEDTSLVKLETTGTGTPSKRELP 1007
QY 781 LLKGSNPSSVSKEKIN 796
Db 1008 TPSWTRDSSLIKETTN 1023

RESULT 19
US-10-270-085-6
; Sequence 6, Application US/10270085
; Patent No. 6627408
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/03/723,820
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 05/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-10-270-085-6

Query Match 3.5%; Score 145.5; DB 4; Length 1073;
Best Local Similarity 18.6%; Pred. No. 0.02;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;
QY 23 GKQATQKTEMLTPEVSKRGINAE-----QIVIKI--TDQGYVTSHGDIHYHYN 70
Db 147 GQTGKTGYTMSGD-LSDSDGILSEAGLIPRALYQLFSSLDNSNQEVAVK-CSYIELYN 204
QY 71 GKPYDAITSEELLKMPNKLKDED-----IVNEVKGGVYKVDGKYVYVVKDAHADN 125
Db 205 EEI-RDLVSEEL--RKPARVFEDTSRGNVITGIEESY-INKAGDGLRLREGSHRRQ 260
QY 126 VRTKEEINRQKQHS-----QHRGGTTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDT 180
Db 261 VAATKCNLDSSRSHSIPTITLHRKVSSGMDTNSLTINN-----NSDDLAR--307
QY 181 GDVIVPHGDHYHYIPKNLSASELAABAFLSGRGNLSNRT-----223
Db 308 -----ASKLHMVDLAGSENI--GRSGAENKRARETQMINQSLLTLGRVI 349

QY 224 -----YR-----RONSNDTSRTNWVPSVSNPCTTNTNTSNNNTNSQASQS 264
Db 350 NALVEKAHHPIPYRESKULTRLLQDSLGKTKTSMIVTVS---STNTLEETISTLEYAARA 406
QY 265 NDIDSLKQLYKPLSQRHVESDGLVFPDPAQI-----TSRTARGVAVPHGDHYHFIPIYSQ 319
Db 407 K---SIRNKQNNQLVFRKVLIKDLVLDIERLKNLNAKKNKGVYLAESTYKELMDRVQ 463
QY 320 MSEL-----EERIAIIPRYRSNHWVPSRPEQSPQPTPEPSPQPPAPNLKI-----369
Db 464 NKDLCQEQARKLEVLNLNKS-----SREQLQYVSKSQEHKKEVEALQLVNSSTE 517
QY 370 -----DNSISIVSLQVRKVGEGYVFEKGISRYVFAKDLPSETVKNLESKLSKOSVSH 424
Db 518 LESVKSNEKIKNELVLEIEKRYKYEAKITTTVAIDLSQYYRESKEYIASLEKLDRT 577
QY 425 LTAKKENVAPRDOEFYDKAYNLL-----TEAH 451
Db 578 ERNKEN-----ENFWNLKENLLTMRSFHGSFTDETNGYFTLLDNDFNAMESLELNTHSN 633
QY 452 KALFXNKGKNSDFOALDKILLER-----LNDESTNKEKLVDDLLAFAPIT 496
Db 634 QLLISMTKITTEHFQSLDEALQASRSCAVPNSSLDLIVSELKOSKNSLLDALEHSLQDIS 693
QY 497 -HPERLGKPN--OIEVTEDE-----VR--IAQL 520
Db 694 MSSQKLGNGISSSELIELQKMKESYRQVQELRSLYNLQTHESQKELMYGVNDIDAL 753
QY 521 ADKYTTS--DGYIFDEHDIISDEGDYVTPMHGSHWIGKDSLSDEKVAQAAYTKKGI 578
Db 754 VKTCTTSLNDADII--LSDYISDQSKFESQKQDIIANIGKIVSNFLQESLYTK-----808
QY 579 LPSPDADVANKPTGDSAAAI--YNRVKGKRIPLVR-----LPMVHEHTVEVK 625
Db 809 -----ADILHSHLNDTNSIRKANEIMNRSEEFLENAASQAEIVGANKERIOKTVE--860
QY 626 NGMLIIIPKDH-VHNIKFAFDDHTYKAPNG-----VTLEDLFATIKYVVEHPDERP 676
Db 861 NGSQKLSKSKAHSNSRSMYDCHLALAESQKQGVNLEVTDLRLLOKVK-----E 911
QY 677 HSDNGWGNASEHVLGKDHSEDPNKNFKADEEPVEETPABPEVQVETKEVQAQKBAEV 736
Db 912 HSEDNTEKEHQQLLDLLESVGNNDNL-----IDSIKTPHTELQ-----950
QY 737 LLAKVTDSLK-----ANATETLAGLN-----NLTLOIMDNNSIMAAE-----EKLLA 780
Db 951 ---KITDHLVKGTTSLANHTNELLGLGDESLCNLETTIEDTSLVKLETTGTGTPSKRELP 1007
QY 781 LLKGSNPSSVSKEKIN 796
Db 1008 TPSWTRDSSLIKETTN 1023

RESULT 20
US-08-961-083-118
; Sequence 118, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2

```
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-09-541-782-6

Query Match      3.5%; Score 145.5; DB 3; Length 1073;
Best Local Similarity 18.6%; Pred. No. 0.02;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

QY 23 GKQATOKTENLTPEVSKREGINAE-----QIVIKI--TDQGVVTSBGHDHYHYN 70
Db 147 QGTGKTGYTMSGD-LSDSGILSEGAGLIPALYQLFSSLDNSQYAVK-CSYVELYN 204
QY 71 GKVPYDAIISBELMKDPNYKLBDE-----IVNEVGKGYVIVKDGKYYVYLKDAAHADN 125
Db 205 EEI-RDLLVSEEL--RKPARVFEDTSRRGNVITGIEESY-IGNAGDGLRLREGSHRRQ 260
QY 126 VRTKEEINRQKQHS-----OHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIEDT 180
Db 261 VAATKCNDSLRSRSHSIFITILHRKVSSGTMDETNSLTINN-----NSDDLRL-- 307
QY 181 GDATIVPHGDHYHYIPKNELSASELAFAFLSGRNLNSRT----- 223
Db 308 -----ASKLHMVDLAGSENI--GRSGAENKRARETGMINQSLTLGRVI 349
QY 224 -----YR-----RQNSDNTSRTNWVPSVSNPGTTNTNTNNSNTNSQASQS 264
Db 350 NALVEKAHHIIPYRESKLTLLQDSLGKTKTSMIVTVS---STNTNLEETISTLEYAARA 406
QY 265 NDIDSLKQLYKLPQSQRHVESDGLVFPDPAQI-----TSRTARGVAVPHGDHYHFIPYSQ 319
Db 407 K---SIRNKPQNNQLVFRKVLIKDLVLDIERLKNDLNATRKNGVYLAESTYKELMDRVQ 463
QY 320 MSEL-----EERIARIIPLYRSNHWVPSRPEQSPQTPPEPSPQAPNPKI----- 369
Db 464 NKDLLCQOARKLEVLIDLNVKS-----SRQLOVYSKSNQEHKEVEALQOLVNSSTE 517
QY 518 LESVKSENEKLNELVLEIEKRKKYETNEAKITTVATDLSQYRESKEYIASLVEKLDRT 577
QY 425 LTAKKENVAPRDOEYDKAYNL-----TEAH 451
Db 578 ERNNKEN-----ENPNWNLKFNLLTMLRSFHGSDFTDNGYFTLLDNFNASMEELLNTHSN 633
QY 452 KALFXNKGNSDFQALDKLLER-----LNDESTNKEKLVDDLLAFAPIT 496
Db 634 QLLISMTKITEHFQSLDEALQASRSCAVPNSSLDLIVSELKSKNSLLDALEHSLQDIS 693
QY 497 HPERLGPNS-----QIEYTEDE-----VR--IAQL 520
Db 694 MSSQKLGNGISSELTLOKMKESYRQLVQELRSLYNLQHTHEESQKELMYGVRNDIDAL 753
QY 521 ADKVTTS--DGYIPDEHDIISDEGDAYVTPHMGHSHWIGKDSLKDEKVAQAQYTKGI 578
Db 754 VKTCTISLANDAII--LSYVIDSQSKFESKQODLIANIGKIVSNFLQONESLYTK----- 808
QY 579 LPPSPADVANKAMPTGDSAAAI--YNRVKGKRIPLVR-----LPYMWVHTVRVK 625
Db 809 -----ADILSHLNDTNSIRKANEIMNRSSEFLRNAASQAEIVGANKERIQTVE-- 860
QY 626 NGNLIIPKDH-YHNIKFAMFDHDTYKAPNG-----YTLEDLPATIKYVVEHPDRP 676
Db 861 NGSQLLDSKSAIHSNRSMDYDCHLALAESQKQGVNLEVQTLDRLLQKVK-----E 911
QY 677 HSDGWMGNASEHVLGKHSDPDKNFYKADBPVEETPAEPVQVEIEKVEAQLKEABV 736
Db 912 HSEDNTEKHQOOLLDESLSVGNNDNL-----IDSIKTPHTLEQ----- 950
QY 737 LIAKVTDSSLK-----ANATETLAGLRN-----NLTLQIMDNNSIMAE-----EKLLA 780
```

```
Db 951 ---KITDVLKGTTSANHTNELLGLGDESCLNLETIEDTSLVKLETTGDTSPSKRELPA 1007
QY 781 LLKGNPSSSVSKKIN 796
Db 1008 TPSWTRDSSLIKETT 1023

RESULT 18
US-09-723-820-6
; Sequence 6, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-09-723-820-6

Query Match      3.5%; Score 145.5; DB 4; Length 1073;
Best Local Similarity 18.6%; Pred. No. 0.02;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

QY 23 GKQATOKTENLTPEVSKREGINAE-----QIVIKI--TDQGVVTSBGHDHYHYN 70
Db 147 QGTGKTGYTMSGD-LSDSGILSEGAGLIPALYQLFSSLDNSQYAVK-CSYVELYN 204
QY 71 GKVPYDAIISBELMKDPNYKLBDE-----IVNEVGKGYVIVKDGKYYVYLKDAAHADN 125
Db 205 EEI-RDLLVSEEL--RKPARVFEDTSRRGNVITGIEESY-IGNAGDGLRLREGSHRRQ 260
QY 126 VRTKEEINRQKQHS-----OHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIEDT 180
Db 261 VAATKCNDSLRSRSHSIFITILHRKVSSGTMDETNSLTINN-----NSDDLRL-- 307
QY 181 GDATIVPHGDHYHYIPKNELSASELAFAFLSGRNLNSRT----- 223
Db 308 -----ASKLHMVDLAGSENI--GRSGAENKRARETGMINQSLTLGRVI 349
QY 224 -----YR-----RQNSDNTSRTNWVPSVSNPGTTNTNTNNSNTNSQASQS 264
Db 350 NALVEKAHHIIPYRESKLTLLQDSLGKTKTSMIVTVS---STNTNLEETISTLEYAARA 406
QY 265 NDIDSLKQLYKLPQSQRHVESDGLVFPDPAQI-----TSRTARGVAVPHGDHYHFIPYSQ 319
Db 407 K---SIRNKPQNNQLVFRKVLIKDLVLDIERLKNDLNATRKNGVYLAESTYKELMDRVQ 463
QY 320 MSEL-----EERIARIIPLYRSNHWVPSRPEQSPQTPPEPSPQAPNPKI----- 369
Db 464 NKDLLCQOARKLEVLIDLNVKS-----SRQLOVYSKSNQEHKEVEALQOLVNSSTE 517
QY 370 -----DNSNLSVLOLVKRGVGYVEEKGISRYVFAKDLPSFTYKNELESKLSKQESVSH 424
Db 518 LESVKSENEKLNELVLEIEKRKKYETNEAKITTVATDLSQYRESKEYIASLVEKLDRT 577
QY 425 LTAKKENVAPRDOEYDKAYNL-----TEAH 451
Db 578 ERNNKEN-----ENPNWNLKFNLLTMLRSFHGSDFTDNGYFTLLDNFNASMEELLNTHSN 633
QY 452 KALFXNKGNSDFQALDKLLER-----LNDESTNKEKLVDDLLAFAPIT 496
Db 634 QLLISMTKITEHFQSLDEALQASRSCAVPNSSLDLIVSELKSKNSLLDALEHSLQDIS 693
```


RESULT 12
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent NO. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

| | | | | |
|-----------------------|-------------------|-------------------|-------------|--------------|
| Query Match | 3.9% | Score 163; | DB 4; | Length 3696; |
| Best local Similarity | 18.5%; | Pred. No. 0.0063; | | |
| Matches 183; | Conservative 141; | Mismatches 348; | Indels 316; | Gaps 42; |

| | | |
|----|------|---|
| Qy | 15 | NNRVSYI--DGKQATQKTEMLTPDEVSKREGINAEQIVIKITD-----QGVVTSHGDIYH 67 |
| Db | 1299 | NNQIDGIVSGRQSGINA---ITPDTSTKRNAKN--DIDIKAADKKIKIQRINDATDEIQ 1353 |
| Qy | 68 | YNGKVPDYDAIISBELLMKDPNYKDKEDIINEVKGGYVIVKD-----110 |
| Db | 1354 | EANKEIE-----EAKIEAKONIQRNSTDQVNEAKTNGINKIENITPATVTKSEARQAVQ 1408 |
| Qy | 111 | ---GKYVYVYLKAAHADNVRTKEBIN-----QKQHSQHQREG--TPRNDGAVALA 157 |
| Db | 1409 | NKAMEQINHITQNTPDATNEEKQEAIRNVSaelARVQAIQNAEHTTQGVKTIKDDAITSL 1468 |
| Qy | 158 | RSQGRYTTDGYIFNASDIIEDTGDAIVPHGDHYHYIPKNELSASELAFAELSGRN 217 |
| Db | 1469 | R-----INAOVBEKE-----SARNA 1483 |
| Qy | 218 | LSNSRTYRRQ---NSDMNTS-----RTNWVPVSNPGTNTIN--TSNNSNTNSOASQSD 266 |
| Db | 1484 | IEQKATQOTQFINNNDNATDEEKEVANNLVIAIKQKSLDINSLSNNVDENAKVAGINE 1543 |
| Qy | 267 | IDSLLKOLYKLPUSQRHVSDGLVDPDA-QITSTARGVAVPHGDHYHFTPYQSOMSELEE 325 |
| Db | 1544 | IANVL-----PATAVSKAKKDIQKLAQQALNQIOTHTQTATTEE 1582 |
| Qy | 326 | RIARIPLRYRSH-----WYDPSRPQSPQPTP---355 |
| Db | 1583 | KEAAIOLANQKSNEARTAIQNEHSNNGVAQAKSNGIHEIBLVMPDADHKKSDAKQSIDNKY 1642 |
| Qy | 356 | -EPSPGQPAPNLKIDSNSSLVSOL---VRKVGEGYVPEEKIGIRYVFAKOLPSETVKNLE 412 |
| Db | 1643 | NEQSNLTINTPDATDEBKQKALDKLAKDAGYKVKVDAQTNQOVSDAKTEADITTIHQ 1702 |
| Qy | 413 | SKLSKQESVSHTTUTAKKE-----NVAPROE-----FYDKAYNLL-----447 |
| Db | 1703 | ANVAKKPSARVELDSKPFEDLKRQINATPNATPEBEKQDAIQRLNGKRDENVKLLINQDRDN 1762 |
| Qy | 448 | -TEAHK-----ALFPXNGRNSDFOALDKILLERL-----NDESTNKEKLVDDLL 499 |
| Db | 1763 | EVEQHNKIGIQLEETHANPTRSD--ALQELQTKFISQTELINNNKDATNEEK---DEA 1817 |
| Qy | 490 | AFLAPITHPERL-----GKPNQSIQTEYDEE--VRIAQ-----ADKYTTSQGYIPDEH 535 |
| Db | 1818 | KRLLEISKNTITINQAQTNQVNDKADNGMNEIATIIIPATTIKTDAKTAIDKBAEQOV 1877 |
| Qy | 536 | DIISDEGDVYVTHMGHSHWIGKDSIGDKEKVAQAQVTEKGILPSPDADVKANPTGDS 595 |
| Db | 1878 | TIINGNDA-----TDESKAFARKLIVERAKI-----EAKSNIITNSD 1913 |

QY 596 AAAIYNRVKGEKRIPLVRLPYMVEHTEVVKGNGLIIHPKHQHYHNINIKPAWFDDHTTYKAP-- 653
D5 :
Db 1914 TE---REVNGAKTNGLEKINIOPSTQTKTNAQEINDKAQEIQIINNTPDATEEKKOE 1970
QY 654 -----NGVTLEDL-----FATIKYYVEHPDER---HSNDGWGNAS 686
D5 :
Db 1971 ATNRVNAGLAQAQTNNNAHSTQEVNESKTSNSTATSKSQPNVIKPTAINSLTQERANNQ 2030
QY 687 EHLVGKDKHSDPKNKF-----KADE--EPVEETPAEPVPQVETEKEVA-----Q 730
D5 :
Db 2031 KTLIGDGNATDDEKRAAKQLVTOKLNEQIKTHESQDNQVDNVKQAATAIKLINANA 2090
QY 731 LKEAEVL-----LAKVTDSSLKAN--AT-----ETLAGLRNNL-----TLQIMD 767
D5 :
Db 2091 HKRQDAINILNTLAESKSDIRANQQATTBEKNTAIQSIDDTLAQRNNINGANTNALVD 2150
QY 768 NNSIMAEAEKLALLKGSPSSSKEKI 795
D5 :
Db 2151 EN-LEDGQKQLORIVLSQTQTOAKADI 2177

RESULT 13
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY;
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GNC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 3.8%; Score 159; DB 4; Length 10182;
Best Local Similarity 18.8%; Pred.No.0.064;
Matches 189; Conservative 126; Mismatches 368; Indels 320; Gaps 43;

| | | |
|----|---|--|
| QY | 13 | KENRRYSYIDGKQATOKENTLTPEYSKREGIN-----AEQIVIKITDQGVTSH 62 |
| D5 | : | |
| Db | 5419 | QKGBESLVNGSNTRGEV-----EHLNEAKSLNANKQLRDKVAEKTNVK-QSSDVINDS 5473 |
| QY | 63 | GDRHYHYNGKV-PYDAIISE-----ELLMK----DPNYKLKDEDIVNEVKGYVI 107 |
| D5 | : | |
| Db | 5474 | TEHQRGDYDALQAEANIINEIGNPTLNKSEIEQKLQQLTDAQNALQGSLLHEAKNNAIT 5533 |
| QY | 108 | KVPGKYVYVYLKDAH--ADNVTKEIINROKQEHSHQREGGTPRND-----CAVALAES 159 |
| D5 | : | |
| Db | 5534 | GIN--KLTALNDARQKAIENVAOQTPIAVNQOLTLDRFINPMQALRKVGGQNNVHQ 5591 |
| QY | 160 | QGRYTTDDGYIFNASDIIEDTGAYIVPHGDHYHIYPKNEL--SASELAAAAFLSGRCN 217 |
| D5 | : | |
| Db | 5592 | QSNFYNEDEQPKHYNDSVQAGOTIIDKLQD--PINWKEIEGAINOINTQTALSGENK 5649 |
| QY | 218 | L-----SNSYTRYRQNSDNTSRNWVPSVSNFGTNTNTSNNSNTNSQASQNSNDISLLK 272 |
| D5 | : | |
| Db | 5650 | LHTDQESTNRQIEGLSSNLTAQINAESKDIAVNOAKTRTDVA-----QKLAAAKEINSAMS 5703 |
| QY | 273 | QLYKLPLSQHVBSGLVPDPAQITSTRARGVAVPHGDHYHFIPYSOMSELERIARIIP 332 |
| D5 | : | |
| Db | 5704 | NL-----RDGI-----QNKEDIKRSSAYINADPTKTAYDQALQNAENIINATP 5747 |
| QY | 333 | LYRSNHWWPDSPPEQSPQTPPEPGFGPPAPNLKIDNSSLVSQLVRKVGEGYFEFK 392 |

```

Qy 244 PGTTNTNNTNSNSQASQNDIDSLIKQYKPLPSQRHVESGLVDFPAQITSRARG 303
Db 225 ---NTQSVAKGTSKPAKSENILQSLKELYDSPAQRYSQESGLVDFPAKIIISRTNG 280
Qy 304 VAVPHGDHYHIFIPYSOMSELEERTARIIPLYRSNHHWVDSRPPQSPQPTPEPSGPQP 363
Db 281 VAIIPHGDHYHIFIPYSKLSALEEKLARVP----- 309
Qy 364 APNLKIDNSGLVQLRVKVGEGYVPEEKGISRYVFAKDLPSQTVKNLESKLSQESVSH 423
Db 310 ---ISGTGVSTNAK-----PNEVVSLSGLSSNPSS--- 339
Qy 424 TLTAKENVAPRQDEFYDKAYNLTAEHAKALFXNKGNSDFQALDKLLERLNDESTNKEK 483
Db 340 ---LTTSKELSSASDGYIFNPK-DIVEETATAYIVRHG--DHPHYTPK-----SNQIGOPT 389
Qy 484 LVDDLAFAPITHPERLKGPNKSOIEVTEDEVRIAQLADKYTTSDDGYIFDEHDIISDEGD 543
Db 390 LPNNSLATPSP-SUPIPFGTSHEKHE-----EDGYGFDFANRIIADES 431
Qy 544 AYVTPHMGHSHWICK 558
Db 432 GFVMSHGDHNYFEK 446

```

RESULT 11

```

US-08-956-171E-5235
; Sequence 5235, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248PL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 5235:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 886 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5235:
US-08-956-171E-5235

Query Match
Best Local Similarity 4.1%; Score 169; DB 4; Length 886;
Matches 179; Conservative 138; Mismatches 343; Indels 246; Gaps 46;

Qy 21 IDGQATQKTENL-TPDEVSKREGINAEQI-VIKITDQGVVTHSGDHYHYHYNKVPYDAI 78
Db 46 ITSQHOQAAAENTNTSDKISENQNNTATTQPPKDTNQTQATQP-----ANTAKNYP- 99
Qy 79 ISELL-----MKDNYKLDKEDI VNEKGGYVI---KVDGKYVYV-LKDAAHADNVTKE 130
Db 100 -ADESLKDAIKPALENKEHDIGPREQVNFQLLDKNNETQYHFFSKDPADVITYYTKKA 158
Qy 131 EI-----NRQK-----QEHQHQREGGTPRNDGAVALARSOQ 161
Db 159 EVELDINTASTWKKFEVYENNQKLPVRLVSYPVPEDHAVIR---FPVSDGTQEL-KIVS 214
Qy 162 RYTTDDGY-----IFNASDIIE-DTGDAYIVPHGDHYHYIIPKNELASELA 206
Db 215 STQIDGDEETNYDTKLVPFAKPIYNDPSLVKSDTNDAAVT--NDQSSVASNOTNTSN 272
Qy 207 AAEAFLSGRGNLSNSRTYRRQNSDNTSRTNWPSVSNPG-TTNTNTNNSNTNSQASQSN 265
Db 273 QNISTINNANNQPAATTNMSQPAQPKSSTNADQASQPAHETNSGNTNDKTNESSQSD 332
Qy 266 DIDLKQLYKPLPSQRHVESDGLVFD---PAQITSRARGVAVPHGDHYHFFPYQSMS 321
Db 333 -----VNOQYPPADESLQDAIKNPAIIDKEHT-----ADNWRPIDFQMKN 372
Qy 322 ELBERIARIIPLYRSNHHWVDSRPEQSPQPTPEPS-----PGQOPANLK----- 368
Db 373 DKGER-----QFYHYA-----STVEPATVIFTKGPILGLGTASTWKK 412
Qy 369 ---IDNSSLSVQLVRKVG-----GYV-FEEKIGISRYVFAKDLPSQTVKNLESKLSQESV 421
Db 413 FEVYEGDKLPVELSVSDSKDYAIRFPVSNGTREV-----KIVSSIEYGENIHEDY 465
Qy 422 SHTLTAKKENVAPRQDEFYD-KAYN---LITEAHKALFXNKGNSDFQALDKLLERLNDE 477
Db 466 DYTLMWFAQPIITNPPDDYVDEETYNLQKLLAPYHKA---KTLERQVVELEKLEKLPK 521
Qy 478 -SINKEKLVDDLAFIA-----PITHEPRLGKNSQIEYTEDEVRIAQLADKYTTS--DG 529
Db 522 YKAEYKKKLDQTRVELADQVKSATFENVTPTNDQDLTDLQEAHFVFESEENSESVMDG 581
Qy 530 YIFDEHDIISD--EGDAYVT-PMHGSHWICKDLSDEKVAQAQAYTKRG-----ILPPS 582
Db 582 FV--EHPFYATLNGQKYVVMKTKDQSYW--KDLIVEGKVTTVTSKPKNSRTLIFPI 637
Qy 583 PDADVKNPTGDSAAAIYNRVKGKRIPLVRLPMVEHTVEVKNGNLIIPKHQHYNI-- 640
Db 638 PD-----KAVYNAI---VKVVVANIGEGQYHVRINIQDINTKDDTSDQNTS 682
Qy 641 -----KEAFWD--DHTYKAPNGYTLDELPAIKYVVEHPDERPHSN---DCWGN 684
Db 683 EPLNVQTGQSGKVADTDVAENSSATNPKNQASDKADVIE-----PSDVVVKDADNN 733
Qy 685 ASEHVLGKDHSEDPNKNFKADEPEVPEETPAEPEVPQVETEKVEAQLEAEVLLAKYTDS 744
Db 734 IDKDVQHDVDHLSDSMDNNHFD-----KYDLKEMDTQIAKDTDR 772
Qy 745 SLKANATETIAGLRNNLT-----QIMDNNSIMAEKLLALLKG-SNPS 788
Db 773 NVDKDADNSV-GMSSNVDTKDSNKNKDKVITQLNHIADKNNHTGKAALKLVVKNQYNTD 831
Qy 789 SVSKEK 794
Db 832 KVTDKK 837

```

```

/ APPLICANT: Choi et. al.
/ TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
/ NUMBER OF SEQUENCES: 452
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII Text
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/961,083
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Brookes, A. Anders
/ REGISTRATION NUMBER: 36,373
/ REFERENCE/DOCKET NUMBER: PB340P2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 182:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 447 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-961-083-182

```

```

Query Match      28.5%; Score 1187.5; DB 3; Length 447;
Best Local Similarity 46.3%; Pred. No. 4.7e-84;
Matches 257; Conservative 72; Mismatches 117; Indels 109; Gaps 10;

QY 4 LGLYQARTVKNRRVSYIDGKQATQKTNLTPEVSKREGINAEQIVIKITDQGYVTSHG 63
DB 1 LQHRQENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIQAEQIVIKITDQGYVTSHG 60
QY 64 DHYHYNGKVPDYAIISEELLMKDPNYKLKDEIVNEVGKGYIKVDGKYYVYLKDAHA 123
DB 61 DHYHYNGKVPDYALFSEELLMKDPNYQLKDADIVNEVGKGYIKVDGKYYVYLKDAHA 120
QY 124 DNVRTKEENRQKQHSQHREGTTPNDGVALARSQGRYTTDDGYIFNASDIIEDTGA 183
DB 121 DNVRTKDEINRQKQHVKNDE---KVNSNVAVARSQGRYTTNDGYVFNPAIDIEDTGA 176
QY 184 YIVPHGDHYYIPKQNELSASELAABAFISGRGNLSNRTYRRQNSDNTSRTNWVPSVSN 243
DB 177 YIVPHGGHYHYIPKSDLSASELAABAFISGRGNLSNRTYRRQNSDNTSRTNWVPSVSN 243
QY 244 PGTTNTNNTNNTNSQAQSQSDIDSLLKQYKLPQSQRHVSDGLVFPDPAITSTARG 303
DB 225 ----NQTSVAKGSTKPAKNSQENLQSLKELYDSPAQRYSESDGLVFPDPAKISRTPNG 280
QY 304 VAVPHGDHYYIPYSQMSSELEIRIARIIPLYRSNHHVDPDSRPEQSPQTPPEPSPGP 363
DB 281 VAIHGDHYYIPYSKLSALEKIAHWV-----PNEVSSLSGLSNPSS--- 309
QY 364 APNLKIDNSSLVQLRVKVGEGYVPEEKIGRYVFAKDPSETVKNLSKLSKQBSVSH 423
DB 310 ----ISGTGSTVSTNAK-----PNEVSSLSGLSNPSS--- 339
QY 424 TLTAKEVNAVDPQEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTNKEK 483
DB 340 -LTTSKELSSASDGYIFNPK-DIVERETATAYIVRHG--DHFHYIPK-----SNQIGQPT 389

```

```

QY 484 LVDDLLAFAPITHPERLKGPNQSIQYETDEVERIAQLADKYTSDGYIFDEHDIISDEGD 543
DB 390 LPNNSLATPSP-SLPINPGTSHKHE-----EDGYGFDPANRIIAEDS 431
QY 544 AYVTPHMGHSHWICK 558
DB 432 GFVMSHGDHNVFFK 446

RESULT 10
US-09-536-784-182
/ Sequence 182, Application US/09536784
/ Patent No. 6573082
/ GENERAL INFORMATION:
/ APPLICANT: Choi et. al.
/ TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
/ NUMBER OF SEQUENCES: 452
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII Text
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/536,784
/ FILING DATE: 30-Oct-1997
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/961,083
/ FILING DATE: OCT-30-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Michelle S. Marks
/ REGISTRATION NUMBER: 41,971
/ REFERENCE/DOCKET NUMBER: PB340P3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 182:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 447 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-536-784-182

```

```

Query Match      28.5%; Score 1187.5; DB 4; Length 447;
Best Local Similarity 46.3%; Pred. No. 4.7e-84;
Matches 257; Conservative 72; Mismatches 117; Indels 109; Gaps 10;

QY 4 LGLYQARTVKNRRVSYIDGKQATQKTNLTPEVSKREGINAEQIVIKITDQGYVTSHG 63
DB 1 LQHRQENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIQAEQIVIKITDQGYVTSHG 60
QY 64 DHYHYNGKVPDYAIISEELLMKDPNYKLKDEIVNEVGKGYIKVDGKYYVYLKDAHA 123
DB 61 DHYHYNGKVPDYALFSEELLMKDPNYQLKDADIVNEVGKGYIKVDGKYYVYLKDAHA 120
QY 124 DNVRTKEENRQKQHSQHREGTTPNDGVALARSQGRYTTDDGYIFNASDIIEDTGA 183
DB 121 DNVRTKDEINRQKQHVKNDE---KVNSNVAVARSQGRYTTNDGYVFNPAIDIEDTGA 176
QY 184 YIVPHGDHYYIPKQNELSASELAABAFISGRGNLSNRTYRRQNSDNTSRTNWVPSVSN 243
DB 177 YIVPHGGHYHYIPKSDLSASELAABAFISGRGNLSNRTYRRQNSDNTSRTNWVPSVSN 243

```


NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-536-784-66

Query Match 63.6%; Score 2649.5; DB 4; Length 763;
Best Local Similarity 66.8%; Pred. No. 2.3e-197;
Matches 519; Conservative 77; Mismatches 132; Indels 49; Gaps 8;
QY 1 SYELGLQARTV-KENNRVSYIDGKOATQKTENLTPDEVSKREGINAEQIVIKITDQGYV 59
Db 2 SYELGRHQAGVKKESNRVSYIDGQAGKAENLTPDEVSKREGINAEQIVIKITDQGYV 61
QY 60 TSHGDHYHYNGKVPYDAIISSELLMKDPNPKLDEDIVNEVKGKGYVVKDGYVYVYVKD 119
Db 62 TSHGDHYHYNGKVPYDAIISSELLMKDPNPKLDEDIVNEVKGKGYVVKDGYVYVYVKD 121
QY 120 AAHADNVRTKEEINRQKQEHSHREGGTPRNDGAVALARSGRYTTDGYIFNADSIIED 179
Db 122 AAHADNVRTKEEINRQKQEHSHREGGTPRNDGAVALARSGRYTTDGYIFNADSIIED 178
QY 180 TGDYIVPHGDHYHYIPKNEISASLAAAEAFSLGRGNLSNRYRRQNSONTSETNWVP 239
Db 179 TGDYIVPHGDHYHYIPKNEISASLAAAEAFSLGRGNLSNRYRRQNSONTSETNWVP 227
QY 240 SVSNPGTTNTNTSNNSNTNSQASNDIDSLIKOLYKPLSORHVESDGLVDPQAITSR 299
Db 228 NPAQRLSENHLITWPIYHQ-NQGENISSLLRELYAKPLSERHVESDGLVDPQAITSR 286
QY 300 TARGVAVPHGDHYHYIPYQMSLEERARIPRYRSHNHWVPSRPPSPQPTPEPSP 359
Db 287 TARGVAVPHGDHYHYIPYQMSLEERARIPRYRSHNHWVPSRPPSPQPTPEPSP 346
QY 360 GPQAPNLK-IDSN---SSLSYQLVRKGVGVEEKGISRYVFAKDLPSRTVKNLESKL 415
Db 347 SPQAPNPQAPSPNIDKLVKEAVRKVGDDGVFEENGVSRYIPAKDLISAETAAGIDSKL 406
QY 416 SKQESVSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKILLERLN 475
Db 407 AKQESLSHKLGAKKTDLPSSDRFYNKAYDILLARHQDLLNKGQVDFEALDNLLELKL 466
QY 476 DESTNKEKLVDDLLAFAPITHPERLGPNSQIETEVDEVIRIAQLADKYTTSDGYIFDEH 535
Db 467 DVXSDKVLXVDLILAFAPIRHPRGLRPNQAITYTDDIEIQVAKLAGYTTEDGYIFDPR 526
QY 536 DIISDEGDYVYTBHMGHSHWGKSLDKVAAQYTKKGLPSPDADVKANPTGDS 595
Db 527 DITSDEGDYVYTBHMGHSHWGKSLDKVAAQYTKKGLPSPDADVKANPTGDS 586
QY 596 AAATYNRVKGKRIPLVRPLVYVHEVTVKNGNLIIPKDHYNHNIKFAWFDHHTYKAPNG 655
Db 587 ABATYNRVKAARKVPLDRMPYNLYTVKNGSLIIPHYDHYHNIKFEWFDGLYEAPKG 646
QY 656 YTLDELFAITYKYVHEPDRPHSNDGWNAGSEHVIGKKDHSEDPKNKFADEE-----P 709
Db 647 YTLDELFAITYKYVHEPDRPHSNDGWNAGSEHVIGKKDHSEDPKNKFADEE-----P 706
QY 710 VEET-----PAPEPQVETEKVQAQLKEAEVILAKVTD 743
Db 707 EETPREKPOSEKPESEPKTEEPSEPESEEPQVETEKVQAQLKEAEVILAKVTD 763

RESULT 8
US-09-468-656A-6
; Sequence 6, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-6
Query Match 28.9%; Score 1203.5; DB 4; Length 484;
Best Local Similarity 45.8%; Pred. No. 3e-85;
Matches 259; Conservative 77; Mismatches 120; Indels 109; Gaps 10;
QY 1 SYELGLQARTV-KENNRVSYIDGKOATQKTENLTPDEVSKREGINAEQIVIKITDQGYV 60
Db 22 AYALNQHRQSEKDNRRVSYVDGSSQSKSENLTPDVQSKGIEQIAEQIVIKITDQGYV 81
QY 61 SHGDHYHYNGKVPYDAIISSELLMKDPNPKLDEDIVNEVKGKGYVVKDGYVYVYVKD 120
Db 82 SHGDHYHYNGKVPYDAIISSELLMKDPNPKLDEDIVNEVKGKGYVVKDGYVYVYVKD 141
QY 121 AAHADNVRTKEEINRQKQEHSHREGGTPRNDGAVALARSGRYTTDGYIFNADSIIED 180
Db 142 AAHADNVRTKEEINRQKQEHSHREGGTPRNDGAVALARSGRYTTDGYIFNADSIIED 197
QY 181 GDYIVPHGDHYHYIPKNEISASLAAAEAFSLGRGNLSNRYRRQNSONTSETNWVP 240
Db 198 GNAYIVPHGDHYHYIPKSDLSASLAAAKAHLAKGNQPSQLSYSSTASDN----- 248
QY 241 VSNPGTTNTNTSNNSNTNSQASNDIDSLIKOLYKPLSORHVESDGLVDPQAITSR 300
Db 249 -----NTQSVAKGSTKPKANSKSENIQSLKELYDSPAQRVSESDGLVDPQAITSR 301
QY 301 ARGVAVPHGDHYHYIPYQMSLEERARIPRYRSHNHWVPSRPPSPQPTPEPSPG 360
Db 302 PNGVAIPHGDHYHYIPYKSLSALAEKARVP----- 333
QY 361 PQAPNLKIDSNLSLVQLVRKGVGVEEKGISRYVFAKDLPSRTVKNLESKLKQES 420
Db 334 -----ISGTGSTVSTNAK-----PNEVVSLSGLSSNPSS 363
QY 421 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKILLERLNDESN 480
Db 364 -----LTSKELSSASDGYIFNPK-DIVEETATAYIVRHG---DHFHYIPK-----SNQIG 410
QY 481 KEKLVDDLLAFAPITHPERLGPNSQIETEVDEVIRIAQLADKYTTSDGYIFDEHDIISD 540
Db 411 QPTLPNNLSLATPSP-SLPINPGTSHEKHE-----EDGYGFDFANRIIAE 452
QY 541 EGDYVYTBHMGHSHWGKSLDKV 565
Db 453 DESGFVMSHGHNHYFFKDLITEQ 477

RESULT 9
US-08-961-083-182
; Sequence 182, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:

QY 596 AAIYNRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHDKHYHNKIFAWPDDHTYKAPNG 655
DB 607 AEAIYNRVKAAKVPDLRMPYNQYIVVEKNGSLIIPHIDHYHNKIFWDFEGLYEAPKG 666
QY 656 YTLDELFAITKYYVEHPDERPHSDNGWGNASEHVLGK-----KDHSE----- 697
DB 667 YTLDELFAITKYYVEHPDERPHSDNGWGNASEHVLGK-----KDHSE----- 726
QY 698 -----DPNKNFKADEPVEZETPAEPVPOVETEKVEAQLKEARVLL 738
DB 727 ESDEKENHAGLNPADNLYKPSDTTETETEEAEDTTDEAEPQVENSVINAKIADAEALL 786
QY 739 AKVTDSLLKANATETLAGLRNNLTQIMDNNSIWAEEKLLALLKGSNPSV 790
DB 787 EKVTDPSIRQANETITGLKSSLLLTGKNNNTISAEVDSLLALLKESQAPI 838

RESULT 6
US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-66

Query Match 63.6%; Score 2649.5; DB 3; Length 763;
Best Local Similarity 66.8%; Pred. No. 2.3e-197;
Matches 519; Conservative 77; Mismatches 132; Indels 49; Gaps 8;
QY 1 SYELGYYQARTV-KENNRVSYIDGKQATKNTLTPDEVSKREGINAEQVIKITDQYV 59
DB 2 SYELGRHQAGQVKESNRVSYIDGQAGKAENLTPEVSKREGINAEQVXVITDQYV 61
QY 60 TSGHDHYHYNGKVPYDAITSEELMKDIPYKLDKEDIVNEVGGYVYKVDGKYVYVKD 119
DB 62 TSGHDHYHYNGKVPYDAITSEELMKDIPYKLDKEDIVNEVGGYVYKVDGKYVYVKD 121

QY 120 AAHADNVRKEBINRQKQBSHQHREGTTPRNDGAVALARSQGRYTTDDGYIFNASDIIE 179
DB 122 AAHADNIRIKKEIKRQKQESHNHNS---RADNAVAARAAQGRYTTDDGYIFNASDIIE 178
QY 180 TGDAYIVPGDGHVHYIPKNELSASELAAPAFUSGRGNLSNRTYRQNSDNTSRTNWVP 239
DB 179 TGDAYIVPGDGHVHYIPKNELSASELAAPAFUSGRGNLSNRTYRQNSDNTSRTNWVP 227
QY 240 SVSNPGTWTNTNNTNSQASQNDIDSLLQLVYKPLSORHVESDGLVFPDPAQITSR 299
DB 228 NPAPQLSENHNLTUPTVTHQ--NQENISLLRELYAKPLSERHVESDGLVFPDPAQITSR 286
QY 300 TARGVAVPHGDHVFHPIYQMSSELEBERIARIIPLYRSNHWDPDSRPEQSPQTPPEPSP 359
DB 287 TARGVAVPHGNVHVFPIYEQMSELEKRIARIIPLYRSNHWDPDSRPEQSPQTPPEPSP 346
QY 360 GPQAPNLK-IDSN-----SSLVLSOLVRKVGEGYFEEKISRYVFAKDLPLSETVKNLESKL 415
DB 347 SPQAPNPQAPSNPIDEKLVKEAVRKVGEGYFEEKISRYVFAKDLPLSETVKNLESKL 406
QY 416 SKOESVSHTLTAKKNVAPRDOEFYDKAYNLLTEAHKALFXNKRNSDFOALDKLLERLN 475
DB 407 AKOESLSHLGAKKTDLPSSDREFYNKAYDILLARIHQDLNDKGRQVDFALDNLLERLK 466
QY 476 DESTNKEKLVDDLLAFLAPITHPERLGKPNQSIQYETEDVIRIAQLADKYTTSDGYIFDEH 535
DB 467 DVKSDKVLVXDILAFAPIRHPERLGKPNQSIQYETEDVIRIAQLADKYTTSDGYIFDPR 526
QY 536 DIISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKKGLPLSPDADVKANPTGDS 595
DB 527 DITSDEGDAYVTPHMTSHWIKDLSLSEAEARAAQAYAKEKGLTPSTDRQDSGNTBAKG 586
QY 596 AAIYNRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHDKHYHNKIFAWPDDHTYKAPNG 655
DB 587 AEAIYNRVKAAKVPDLRMPYNQYIVVEKNGSLIIPHIDHYHNKIFWDFEGLYEAPKG 646
QY 656 YTLDELFAITKYYVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNFKADEE-----P 709
DB 647 YTLDELFAITKYYVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNFKADEE-----P 706
QY 710 VEET-----PAEPVPOVETEKVEAQLKEARVLLAKVTD 743
DB 707 EETPREKQSEKPSKPTPEEPSESEBPOVETEKVEAQLKEARVLLAKVTD 763

RESULT 7
US-09-536-784-66
; Sequence 66, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:

| | | | |
|---|-----|--|-----|
| Db | 724 | PREKQSEKPEPKPTPEPSEBPQVETEKVEKLEAREADLLKIQDPIKSN | 785 |
| Qy | 751 | TETLAGLRNNLTQIMDNNSIMAAEKLLALLKGS | 785 |
| Db | 784 | KEITLGLKNNLLPGTODNNTIMAAEKLLALLKES | 818 |
| <p>RESULT 5</p> <p>US-09-468-656A-4</p> <p>Sequence 4, Application US/09468656A</p> <p>Patent No. 6582706</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Johnson, Leslie S.</p> <p>APPLICANT: Adamou, John E.</p> <p>TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus</p> <p>TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural</p> <p>TITLE OF INVENTION: Motifs</p> <p>FILE REFERENCE: 469201-444</p> <p>CURRENT APPLICATION NUMBER: US/09/468,656A</p> <p>CURRENT FILING DATE: 1999-12-02</p> <p>PRIOR FILING DATE: 1998-12-21</p> <p>PRIOR FILING DATE: 1998-12-21</p> <p>NUMBER OF SEQ ID NOS: 14</p> <p>SOFTWARE: PatentIn Ver. 2.1</p> <p>SEQ ID NO 4</p> <p>LENGTH: 838</p> <p>TYPE: PRT</p> <p>ORGANISM: Streptococcus pneumoniae</p> <p>US-09-468-656A-4</p> | | | |
| <p>Query Match 66.6%; Score 2772; DB 4; Length 838;</p> <p>Best Local Similarity 65.0%; Pred. No. 7.8e-207;</p> <p>Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8</p> | | | |
| Qy | 1 | SYELGLYQARTV-KENNRVSYIDGQAQTKTENLTPDEVSKREGINAEQIVIKITDQGYV | 59 |
| Db | 21 | SYELGRHQAGQKESNRVSYIDGQAQKAENLTPDEVSKREGINAEQIVIKITDQGYV | 80 |
| Qy | 60 | TSBGHDHYHYNGKVPYDAIISSELLMKDPNYKLKDEDIWNEVKGGYVIVKDGKYYVYLKD | 119 |
| Db | 81 | TSBGHDHYHYNGKVPYDAIISSELLMKDPNYQLKDSIWIWELKGGYIVKNGKYYVYLKD | 140 |
| Qy | 120 | AAHADNVRTKEEINRQKQHSQREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIED | 179 |
| Db | 141 | AAHADNVRTKEEIKRQKQSHNHGGS--NQAVVAARAQGRYTTDDGYIFNASDIIED | 198 |
| Qy | 180 | TGDAYIVPHGDHYHYIPKNELSASELAARAEFLSGRGLNSRTYRRQNSDNTSRTNWVP | 239 |
| Db | 199 | TGDAYIVPHGDHYHYIPKNELSASELAARAEFLSGRGLNSRTYRRQNSDNTSRTNWVP | 247 |
| Qy | 240 | SVSNPGTNTNTSNNSNTNSQASQNDIDSLKQLYKLPLSRHVESDGLVDPQAITSR | 299 |
| Db | 248 | NPAQPLSENHNLTVPTTHQ--NQENISLLRELYAKPLSERHVESDGLIFDPAQITSR | 306 |
| Qy | 300 | TARGVAVPHGDHYHYIPYQMSSELEERIIARIPLRYRSHNHVPSRPEQSPQPTPEPSP | 359 |
| Db | 307 | TARGVAVPHGNHYHYIPYEQMSELEKRIARIPLRYRSHNHVPSRPEQSPQSTPEPSP | 366 |
| Qy | 360 | GPQAPNLK-IDSN---SSLSQLVKVGEGYVFEKGLSRVFAKDLPSFTVKNLESKL | 415 |
| Db | 367 | SPQAPNPAPSNPIDEKLVKEAVRKVGDDGYVFEENGVSRYIPAKDLSAETAAGIDSKL | 426 |
| Qy | 416 | SKQESYSHLTAKENVAPRDOEFYDKAYNLLTEAHKALFKXNKGNSDFOALDKLLERLNL | 475 |
| Db | 427 | AKQESLHKLGAKKTDLPSSDREFYNKAYDILLARLHQDLLDNKGRQVDFEALDNLERLK | 486 |
| Qy | 476 | DESTNKEKLVDDLAFAPITHPERLKGPNQSIETDEVRITQAQADKYTTSDGYIFDEH | 535 |
| Db | 487 | DVPSDKVLVDLAFAPIRPERLKGPNQAQITTYDDEIQVAKLAGKTYTDEDGYIFDPR | 546 |
| Qy | 536 | DIISDEGDYVTPHMGHSHWIKDLSDEKKEVAAQAYTKEGILPSPDPAVKANPTGDS | 595 |
| Db | 547 | DIISDEGDYVTPHMGHSHWIKDLSDEKKEVAAQAYTKEGILPSPDPAVKANPTGDS | 606 |

```

QY 1 SYELGLYQARTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDOGYVT 60
DB 1 SYELGLYQARTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDOGYVT 60
QY 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDEDIVNEVKGYYIKVDGKYVYLKDA 120
DB 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDEDIVNEVKGYYIKVDGKYVYLKDA 120
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGVALARSGRYTTDDGYIFNASDIIEDT 180
DB 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGVALARSGRYTTDDGYIFNASDIIEDT 180
QY 181 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
DB 181 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
QY 241 VSNPGTTNTNTSNNSTNSQASQSDNIDSLKQYKPLPSQRHVESDGLVDFDPAQITST 300
DB 241 VSNPGTTNTNTSNNSTNSQASQSDNIDSLKQYKPLPSQRHVESDGLVDFDPAQITST 300
QY 301 ARGVAVPHGDHYHYIIPYSQMSSELEERLARIIPLYRSNHNWVPSRPPQPTPEPSPG 360
DB 301 ARGVAVPHGDHYHYIIPYSQMSSELEERLARIIPLYRSNHNWVPSRPPQPTPEPSPG 360
QY 361 PQAPNLIKIDNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSKVKNLESKLSKQES 420
DB 361 PQAPNLIKIDNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSKVKNLESKLSKQES 420
QY 421 VSHTLTAKKENVAPRQDFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
DB 421 VSHTLTAKKENVAPRQDFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
QY 481 KEKLVDDLLAFLAPITHPERLGPNSQIETEDVRIQAQADKYTTSDGYIFDEHDIISD 540
DB 481 KEKLVDDLLAFLAPITHPERLGPNSQIETEDVRIQAQADKYTTSDGYIFDEHDIISD 540
QY 541 EGDAYVTPHMGHSHWIGKSDLSKDKVAAQAYTKEGILPPSPDADVKANPTGDSAAIY 600
DB 541 EGDAYVTPHMGHSHWIGKSDLSKDKVAAQAYTKEGILPPSPDADVKANPTGDSAAIY 600
QY 601 NRKVGKRIPLVRLPYMVEHTEVKGNNLIIPKHQHYHNKFAWDFDHTTKAPNGYTTLED 660
DB 601 NRKVGKRIPLVRLPYMVEHTEVKGNNLIIPKHQHYHNKFAWDFDHTTKAPNGYTTLED 660
QY 661 LFATIKYVVEHPDERPHSDNGWGNASHEVLGKXKHSDDPNKFKADEEPVEETPAEPEVP 720
DB 661 LFATIKYVVEHPDERPHSDNGWGNASHEVLGKXKHSDDPNKFKADEEPVEETPAEPEVP 720
QY 721 QVETEKVEAQLKEAEVLLAKVTDSLLKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
DB 721 QVETEKVEAQLKEAEVLLAKVTDSLLKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
QY 781 LLKGSNPSSVSKKIN 796
DB 781 LLKGSNPSSVSKKIN 796

```

```

RESULT 3
US-09-468-656A-8
; Sequence 8, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21

```

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-8

Query Match 100.0%; Score 4163; DB 4; Length 819;
Best Local Similarity 99.9%; Pred. No. 5.1e-315; Indels 0; Gaps 0;
Matches 795; Conservative 0; Mismatches 1;

QY 1 SYELGLYQARTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDOGYVT 60
DB 21 SYELGLYQARTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDOGYVT 80
QY 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDEDIVNEVKGYYIKVDGKYVYLKDA 120
DB 81 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDEDIVNEVKGYYIKVDGKYVYLKDA 140
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGVALARSGRYTTDDGYIFNASDIIEDT 180
DB 141 AHADNVRTKEEINRQKQHSQHREGGTPRNDGVALARSGRYTTDDGYIFNASDIIEDT 200
QY 181 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
DB 201 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 260
QY 241 VSNPGTTNTNTSNNSTNSQASQSDNIDSLKQYKPLPSQRHVESDGLVDFDPAQITST 300
DB 261 VSNPGTTNTNTSNNSTNSQASQSDNIDSLKQYKPLPSQRHVESDGLVDFDPAQITST 320
QY 301 ARGVAVPHGDHYHYIIPYSQMSSELEERLARIIPLYRSNHNWVPSRPPQPTPEPSPG 360
DB 321 ARGVAVPHGDHYHYIIPYSQMSSELEERLARIIPLYRSNHNWVPSRPPQPTPEPSPG 380
QY 361 PQAPNLIKIDNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSKVKNLESKLSKQES 420
DB 381 PQAPNLIKIDNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSKVKNLESKLSKQES 440
QY 421 VSHTLTAKKENVAPRQDFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
DB 441 VSHTLTAKKENVAPRQDFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
QY 481 KEKLVDDLLAFLAPITHPERLGPNSQIETEDVRIQAQADKYTTSDGYIFDEHDIISD 540
DB 501 KEKLVDDLLAFLAPITHPERLGPNSQIETEDVRIQAQADKYTTSDGYIFDEHDIISD 560
QY 541 EGDAYVTPHMGHSHWIGKSDLSKDKVAAQAYTKEGILPPSPDADVKANPTGDSAAIY 600
DB 561 EGDAYVTPHMGHSHWIGKSDLSKDKVAAQAYTKEGILPPSPDADVKANPTGDSAAIY 620
QY 601 NRKVGKRIPLVRLPYMVEHTEVKGNNLIIPKHQHYHNKFAWDFDHTTKAPNGYTTLED 660
DB 621 NRKVGKRIPLVRLPYMVEHTEVKGNNLIIPKHQHYHNKFAWDFDHTTKAPNGYTTLED 680
QY 661 LFATIKYVVEHPDERPHSDNGWGNASHEVLGKXKHSDDPNKFKADEEPVEETPAEPEVP 720
DB 681 LFATIKYVVEHPDERPHSDNGWGNASHEVLGKXKHSDDPNKFKADEEPVEETPAEPEVP 740
QY 721 QVETEKVEAQLKEAEVLLAKVTDSLLKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
DB 741 QVETEKVEAQLKEAEVLLAKVTDSLLKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 800
QY 781 LLKGSNPSSVSKKIN 796
DB 801 LLKGSNPSSVSKKIN 816

```

```

RESULT 4
US-09-468-656A-10
; Sequence 10, Application US/09468656A
; Patent No. 6582706

```

;; APPLICANT: Choi et. al.
;; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
;; NUMBER OF SEQUENCES: 452
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Human Genome Sciences, Inc.
;; STREET: 9410 Key West Avenue
;; CITY: Rockville
;; STATE: Maryland
;; COUNTRY: USA
;; ZIP: 20850
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;; COMPUTER: HP Vectra 486/33
;; OPERATING SYSTEM: MSDOS version 6.2
;; SOFTWARE: ASCII Text
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/961,083
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Brookes, A. Anders
;; REGISTRATION NUMBER: 36,373
;; REFERENCE/DOCKET NUMBER: PB340P2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (301) 309-8504
;; TELEFAX: (301) 309-8512
;; INFORMATION FOR SEQ ID NO: 56:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 796 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-961-083-56

Query Match 100.0%; Score 4163; DB 3; Length 796;
Best Local Similarity 100.0%; Pred. No. 4.9e-315;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENRRYSYIDGKATQKNTENLTPEVSKREGINAEQIVIKITDQGVVT 60
Db 1 SYELGLYQARTVKENRRYSYIDGKATQKNTENLTPEVSKREGINAEQIVIKITDQGVVT 60
QY 61 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLKDEDI VNEVKGGYVIKVDGKYVYLKDA 120
Db 61 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLKDEDI VNEVKGGYVIKVDGKYVYLKDA 120
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 180
Db 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 180
QY 181 GDAYIVPHGDHYHYTPKNELSASELAAAEAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 240
Db 181 GDAYIVPHGDHYHYTPKNELSASELAAAEAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 240
QY 241 VSNPCTTNTSNNTNSQASQSDNDISLLKQLYKLP LSRHVSDGLVDPDAQITSTRT 300
Db 241 VSNPCTTNTSNNTNSQASQSDNDISLLKQLYKLP LSRHVSDGLVDPDAQITSTRT 300
QY 301 ARGVAVPHGDHYHYTPYSCMSELEERIIITPLRYRSNHWVPDSRPEQSPQPTPEPSPG 360
Db 301 ARGVAVPHGDHYHYTPYSCMSELEERIIITPLRYRSNHWVPDSRPEQSPQPTPEPSPG 360
QY 361 PQAPNLIKIDNSSLSVLSQVLRVKVGGYVFEKIGSYVFAKDLPSGTVKNLESKLSKQES 420
Db 361 PQAPNLIKIDNSSLSVLSQVLRVKVGGYVFEKIGSYVFAKDLPSGTVKNLESKLSKQES 420
QY 421 VSHTLTAKENVAPRQDFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
Db 421 VSHTLTAKENVAPRQDFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480

QY 481 KEKLVDDLLAFAPLADITHPERLGKPNQSIETYTEDEVRIAQLADKYTTSDGYIFDEHDIISD 540
Db 481 KEKLVDDLLAFAPLADITHPERLGKPNQSIETYTEDEVRIAQLADKYTTSDGYIFDEHDIISD 540
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEKGILPPSPDADVKANPTGDSAAAIY 600
Db 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEKGILPPSPDADVKANPTGDSAAAIY 600
QY 601 NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFAWFDDHTYKAPNGYTTLED 660
Db 601 NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFAWFDDHTYKAPNGYTTLED 660
QY 661 LPAIKYVVEHPDPRPHSDNGWGNASEHVLGKDHSEDPNKNFKADBEVPVETPAEPEVP 720
Db 661 LPAIKYVVEHPDPRPHSDNGWGNASEHVLGKDHSEDPNKNFKADBEVPVETPAEPEVP 720
QY 721 QVETEKVEAQLKEAEVLLAKVTTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 780
Db 721 QVETEKVEAQLKEAEVLLAKVTTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 780
QY 781 LLKGSNPSSVSKEKIN 796
Db 781 LLKGSNPSSVSKEKIN 796

RESULT 2
US-09-536-784-56
; Sequence 56, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-536-784-56

Query Match 100.0%; Score 4163; DB 4; Length 796;
Best Local Similarity 100.0%; Pred. No. 4.9e-315;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 652 APNGYTLEDFATIKYVVEHPDE 674
Db 777 -----EHEDE 781

Search completed: October 1, 2004, 07:16:43
Job time : 92 secs

330 KDI-----DLSKEIPNPH-----HDDHEDHKGHHHDSHDH----- 360
461 NSDFQALDKLLERLNDSTNKEKLVDD---LLAFAPATHPERLGKPNQSIYETDEVDRI 517
361 -----HEEHEHTKSNKLSDEQKGLIYLA-----EKLGLNPNQIEV----- 396
518 AQLADKYTSDGYIFDEHDIISDEGDAVVTPHMGHSHW-----IGK-----DLSLD 563
397 -----LTSEDSIIFKY-----PHDHSHTIASKDIKPIPDGHHDSHA 438
564 KEKVAQAAYTKKKG-----IL---PPSPDADVKANP---TGDSAAAIYNRVKGKRIIP 610
439 KDKV-GMATLKQIGFDDBIIOILLHADAPTTPPSNETNPEKMRQWLATVTKINIGQRTNP 497
611 LVRLPYVVEHTVEVKNGLIIPHKD-----HYHNKIFAW-----FDDHTY--KAP----- 653
498 FQREGLSLMPNIEVL-GIGFTPIINDMTPLVQFKLKQLMWTNTGITDYSFLDKMPLLEGL 556
654 ---NGYLEDLFATIKY-----YVEHP-----DERPHS 678
557 DISQNG--IKDLSFLTKYKQLSLTAAANGITSLKPLAELNQLVLVSHNNISDLTPLS 614
679 N-----DGMGNASEHVLGKHSE-DPNKNFKADEEPEETPAE----- 716
615 NLTKLQELHLDHNNVKNLSA-LSGKKDLKVLDSNNKSADLSLTKTTSLETLNNTS 673
717 -----PEVPOV-----ETEKVEAQLKEAEVLLAKVTDSSLK----- 747
674 NLSFLKQNPVSNITINNNAKLASLDGIEESDEIVKVEAEGNQIKSLVKNKQGLKFLNV 733
748 -ANATETIAGLRNNLTQIIMDNNSIMAEAEKLLALLKGNSSPSVSKEKI 795
734 TNNQILSLEGVNNTSLETL-----SVSKNKL 760

RESULT 23
Q8E029
ID Q8E029 PRELIMINARY; PRT; 877 AA.
AC Q8E029;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Protein of unknown function/lipoprotein, putative.
GN SAG0907.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettein H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Bean M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014233; AAM99793.1; -;
DR TIGR; SAG0907; -;
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007092; LRR SDS22.
DR InterPro; IPR006270; Strép_his_triad.
DR Pfam; PF00560; LRR; 3.
DR Pfam; PF04270; strep_his_triad; 4.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 877 AA; 97563 MW; 4829E9E593A4F27 CRC64;

Query Match 7.3%; Score 302; DB 16; Length 877;
Best Local Similarity 20.8%; Pred. No. 1.4e-08;
Matches 200; Conservative 95; Mismatches 263; Indels 402; Gaps 43;
QY 13 KENNRVVIDGKQATOKTENLTDPDEV---SKREGI-----NAEQIVIKITD 55
DB 26 QEHKNSHIKTQVAKKANKKKVSVKESHKKRGVAGVDPTDDGELLTKDSKILSHTD 85
QY 56 QGYVTSGDHYHYNGKVPYDAIISEELLMKDPNKLKDEIVNEVKGYYIKVDGYYV 115
DB 86 SGIVVAHGNHSHFT-----FYSDLKSGKFS 110
QY 116 YLKDAAHADNVRTKEEINROKQHSQHREGTTPNDGAVALARSQRYTTDDGVYENASD 175
DB 111 YL-----IPNANTKTNKQAVRNPKA-----GAVAV-----NTLNDGVVFNPAD 149
QY 176 IIEDTDAYIVPHGDHYHYIPKNELSASELAFAFLSGRGLNSRTYRQNSDNTSRT 235
DB 150 IVSEDAANGYVVRGDFHYIPKASLSQKQVQASEAVSLGNQNS--HYRVNSKSIAGL 207
QY 236 NWPSVSNPGTTNTNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTS 295
DB 208 HY-PT-----SNGLFNGRG 221
QY 296 ITSETARGVAVPHGDHYHYIPYQSOMSELEBRIARIIPLYVRSNHW-----VPDS 344
DB 222 IKGTTPTGILVEHNNHLLHFISFADL-----KKGWGSIAIDRYQPKKADS 266
QY 345 RPEQSPQTPPEPSFGPQPNLKNIDNSSIVSOLVRKV-----GE-GYVF 389
DB 267 KKQSPS-----SKKPTENTLPKIDKDL-AVLARELHLDISIRVLKTLNGEIGFEY 318
QY 390 EEKIGSYVFAKOLPSETVKNLESKLSQBSVSHTLTAKKENVAPROEFVDKAYNLLTE 449
DB 319 PHDDHTHVIWAKDI-----DLSKPIPNPH-----HDDHEDHKGHHHDS 357
QY 450 AHKALFXNKGKNSDFQALDKLLERLNDSTNKEKLVDD---LLAFAPATHPERLGKPN 506
DB 358 DHK-----HEEHEHTKSNKLSDEQKGLIYLA-----EKLGLNPN 392
QY 507 QIYETEDEVRIQALADKYTSDGYIFDEHDIISDEGDAVVTPHMGHSHW-----IGK- 558
DB 393 QIEV-----LTSEDSIIFKY-----PHDHSHTIASKDIKPI 427
QY 559 -----DLSDKKXVAAQYTKKKG-----IL---PPSPDADVKANP---TGDSAAAI 599
DB 428 IPDGHHDHSHAKDKV-GMATLKQIGFDDBIIOILLHADAPTTPPSNETNPEKMRQWLATV 486
QY 600 YNRVKGKRIPLVRLPYVVEHTVEVKNGLIIPHKD-----HYHNKIFAW-----FDDHT 649
DB 487 TKINIGQRTNPFQFGLSLMPNIEVL-GIGFTPIINDMTPLVQFKLKQLMWTNTGITDYS 545
QY 650 Y--KAP-----NGYLEDLFATIKY-----YVEHP----- 672
DB 546 FLDKMPLEGLDISQNG--IKDLSFLTKYKQLSLTAAANGITSLKPLAELNQLVLS 603
QY 673 -----DERPHS-----DGMGNASEHVLGKHSE-DPNKNFKADEEPEETPA 715
DB 604 HNNISDLTPLSLNLTQELYLDHNNVKNLSA-LSGKKDLKVLDSNNKSADLSLTKTTS 662
QY 716 E-----PEVPOV-----ETEKVEAQLKEAEVLLAK 740
DB 663 ETLNLTNETNNTSLNLSFLKQNPVSNITINNNAKLASLDGIEESDEIVKVEAEGNQIKSLVLK 722
QY 741 VTDSSLK-----ANATETIAGLRNNLTQIIMDNNSIMAEAEKLLALLKGNSSPSVSKEKI 795
DB 723 NKQGLKFLNVNNTNQLTSLEGVNNTSLETL-----SVSKNKL 760

RESULT 24
Q8DQ06
ID Q8DQ06 PRELIMINARY; PRT; 182 AA.


```
Db 81 YVTHSGDHVHYFNGKVPYDAIISEELMTDPNHYFKQSDVINELDGYVIKNGNYYVYL 140
QY 118 K0AAHADNVRTKEEINRQKQEHSHQ-REGG-----TPRNDGAVALARSGQRYTTDDGY 169
Db 141 KPGSKRNIRTKQOIAEAVAKGKEAKGLAQVAHLSKEEVAANAEKRGQRYTTDDGY 200
QY 170 IFNASDIIEDTGDAYIVPHGDHVIYIPKNELSASELAARAFILS---GRNLNSRTRYR 226
Db 201 IFSPDTIIDLDGAYIVPHGNHVIYIPKDLSELAARAFILS---GRNLNSRTRYR 258
QY 227 QNSDNTSRNWNVPSVS-NFGTNTNTSN-----NSNTNSQASQND-----IDSLK 272
Db 259 TPAPGRKAP-IPDVTPNPGQHPDNGGYPHPAPPNDASQNKHQDFEFGKTFKELL 317
QY 273 QLYKLPDSORHVESGLVDPQAQITRTARGVAVPHGDHVIYIPYQMSLEBRIARIIP 332
Db 318 QLRHLKLRHVHEEDGLIPEPTQVKSNAFYVPHGDHVIIPRSQLSPELMELAD--- 374
QY 333 LRYRSNHWVPDSRPEQSPQTPPEPSPGPQAPNKLKIDSNSSLSVLSQVRKVGEGYVFEK 392
Db 375 -RYLAG-----QTDNDNS-----GSDH----- 390
QY 393 GISRYVFAKDLPSFTVKNLESKLSQESVSHYTLAKKNVAPRQDFYDKAYNLLTEAHK 452
Db 391 -----SKPSDKE-VTHTFGLHR-----IKAY----- 410
QY 453 ALFXNKGNSDFQALDKLLERLNDESTNKEKLVDLLAFAPITHPERLGKPNQSEIYTE 512
Db 411 -----GKGLD-----GKP----- 418
QY 513 DEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAAY 572
Db 419 -----YDTSAYVFSKESIHSVDKSGVTAKHGDHFIYIGFGELEQYELDEVANW 467
QY 573 TKEKG-----ILPPSPDADVKANPTGDSAAA-----IYNR----- 602
Db 468 VKAQGADELVAALDQEGKEKPLFTTKVSRKVTGDKGVYIMPKDGDYFYARYQLDL 527
QY 603 -----VKGEK-----IPLVRLPYMVEHTVEVKNGLIIPKH 636
Db 528 TQIAFAEQLMLKDKRKHRYDIVDTGIEPLRALVDLSSPLMHAGNATYDTGSSFVIDH 587
QY 637 YHNKFAWFDHHTYKAPNGYTTLEDLPATIKYVVEHDPHPSNDG-----GNASEH 688
Db 588 IHVVPYSWL-----TENQIATIKYVMOHPEVRP---DWMSKPGHESSGVSIPN 632
QY 689 VLKGDHSDPNKFNKADPEPVEETPAE-----PE-----VP 720
Db 633 VTPLDKRAGPNMQIITHSAEVOKALAEGRFAAPDGYIFDPRDLAKETFWVKDGSFIP 692
QY 721 QVE-----TEKVEAQLKEAEVLAKVTDSSSKANATETLAGLNNLTQIMDNNSI 771
Db 693 RADGSLRTRINKSDLSQAEWQQOELLAKN-----AGDATDT-----DKPEE 735
QY 772 MAEAEKLLALLKGSNPFSSVSSEK 794
Db 736 KQQAQK---SNENQQPSEASKER 755
```

RESULT 18

```
Q9ZHG7
ID Q9ZHG7 PRELIMINARY; PRT; 822 AA.
AC Q9ZHG7;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=R268;
RX MEDLINE=99115568; PubMed=9916102;
RA Spellerberg B., Rozdzinski E., Martin S., Weber-Heymann J.,
RA Schmitz N., Iuetzchen R., Podbielski A.;
RT "Lmb, a protein with similarities to the Irai adhesin family, mediates
RL attachment of Streptococcus agalactiae to human laminin.";
RL Infect. Immun. 67:871-878 (1999).
DR EMBL: AF062533; AAD13797.1; -.
DR PIR: T46758; T46758. Strep_his triad.
DR InterPro: IPR006270; Strep_his triad.
DR Pfam: PF04270; Strep_his triad; 6.
DR TIGRFAMs: TIGR01363; Strep_his triad; 4.
KW Hypothetical protein.
SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;
```

Query Match 22.4%; Score 934; DB 2; Length 822;
Best Local Similarity 29.4%; Pred. No. 5.3e-44;
Matches 271; Conservative 99; Mismatches 318; Gaps 32;

```
QY 1 SYELGLYQARTVKENNRVSID--GKQATQKTENLTPEVSKREGINAEQIVIKITDQG 57
Db 22 SYQLGKHMLATKQIAIYIDDSKGVKAPKT-NKTMDOISAEGLISAEQIVVKITDQG 80
QY 58 YVTHSGDHVHYNGKVPYDAIISEELMKDPNPKLDEIVNEVKGGYVIVKDGKYYVYL 117
Db 81 YVTHSGDHVHYNGKVPYDAIISEELMTDPNHYFKQSDVINELDGYVIKNGNYYVYL 140
QY 118 K0AAHADNVRTKEEINRQKQEHSHQ-REGG-----TPRNDGAVALARSGQRYTTDDGY 169
Db 141 KPGSKRNIRTKQOIAEAVAKGKEAKGLAQVAHLSKEEVAANAEKRGQRYTTDDGY 200
QY 170 IFNASDIIEDTGDAYIVPHGDHVIYIPKNELSASELAARAFILS---GRNLNSRTRYR 226
Db 201 IFSPDTIIDLDGAYIVPHGNHVIYIPKDLSELAARAFILS---GRNLNSRTRYR 258
QY 227 QNSDNTSRNWNVPSVS-NFGTNTNTSN-----NSNTNSQASQND-----IDSLK 272
Db 259 TPAPGRKAP-IPDVTPNPGQHPDNGGYPHPAPPNDASQNKHQDFEFGKTFKELL 317
QY 273 QLYKLPDSORHVESGLVDPQAQITRTARGVAVPHGDHVIYIPYQMSLEBRIARIIP 332
Db 318 QLRHLKLRHVHEEDGLIPEPTQVKSNAFYVPHGDHVIIPRSQLSPELMELAD--- 374
QY 333 LRYRSNHWVPDSRPEQSPQTPPEPSPGPQAPNKLKIDSNSSLSVLSQVRKVGEGYVFEK 392
Db 375 -RYLAG-----QTDNDNS-----GSDH----- 390
QY 393 GISRYVFAKDLPSFTVKNLESKLSQESVSHYTLAKKNVAPRQDFYDKAYNLLTEAHK 452
Db 391 -----SKPSDKE-VTHTFGLHR-----IKAY----- 410
QY 453 ALFXNKGNSDFQALDKLLERLNDESTNKEKLVDLLAFAPITHPERLGKPNQSEIYTE 512
Db 411 -----GKGLD-----GKP----- 418
QY 513 DEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAAY 572
Db 419 -----YDTSAYVFSKESIHSVDKSGVTAKHGDHFIYIGFGELEQYELDEVANW 467
QY 573 TKEKG-----ILPPSPDADVKANPTGDSAAA-----IYNR----- 602
Db 468 VKAQGADELVAALDQEGKEKPLFTTKVSRKVTGDKGVYIMPKDGDYFYARYQLDL 527
QY 603 -----VKGEK-----IPLVRLPYMVEHTVEVKNGLIIPKH 636
Db 528 TQIAFAEQLMLKDKRKHRYDIVDTGIEPLRALVDLSSPLMHAGNATYDTGSSFVIDH 587
QY 637 YHNKFAWFDHHTYKAPNGYTTLEDLPATIKYVVEHDPHPSNDG-----GNASEH 688
Db 588 IHVVPYSWL-----TENQIATIKYVMOHPEVRP---DWMSKPGHESSGVSIPN 632
QY 689 VLKGDHSDPNKFNKADPEPVEETPAE-----PE-----VP 720
```

[illegible]

| | | | |
|--|--|--|--------------|
| Qy | 508 | I EYTEDVRIAQLADKYTTSDGYIFDEHDIISDEGDVAVVTPHMGSHHWIGKDSLSDKKY | 566 |
| Db | 494 | -----TKYSRVKTKDGK-----GYMPKDGKDYFYARDQLD-----L | 527 |
| Qy | 568 | AAQAYTKEKGLPPSPADVKANPTGDSAAAIYNRVKGEKRIPLVRLPMVVEHTVEVKNG | 627 |
| Db | 528 | TQIAFAQELMLKDNHRYDIVTG-----IEPLAVDVSSLPWHAGNATYDTGS | 578 |
| Qy | 628 | NLIIPKDHYHNIKPAWPDHDTYKAPNGYTTLEDLFATIKYVVEHPDERPHSNDGM----- | 682 |
| Db | 579 | SFVIPHIDHIIHVFPYSWL-----TRDQIATIKYVMQHPEVRP-----DVMSKPGH | 623 |
| Qy | 683 | -----GNASEHVLGKDHSEDPKNFKADEEPEVETPAE-----PE----- | 718 |
| Db | 624 | EESGSVNPVTPLDKRAQMPNWIHSABEYQKALAEGRFATPDGYIFDPDVLAKETFV | 683 |
| Qy | 719 | -----VPOVE-----TEKVEAQLKEAEVLAKVTTDSSLKANATETLAGRLNLT | 762 |
| Db | 684 | WKDGSFIPRADGSSSLRTINKSDLSQAEWQQAQELLAKN-----AGDATDT----- | 730 |
| Qy | 763 | LQIMDNNSIMAAEKLIALLKGSNPSSVSKEK | 794 |
| Db | 731 | -----DKPKEQQAQK-----SNENQOPSEASKEE | 755 |
| RESULT 17 | | | |
| ID | Q8DZ81 | PRELIMINARY; | PRT; 822 AA. |
| AC | Q8DZ81 | | |
| DT | 01-MAR-2003 | (TREMBLrel. 23, Created) | |
| DT | 01-MAR-2003 | (TREMBLrel. 23, Last sequence update) | |
| DT | 01-OCT-2003 | (TREMBLrel. 25, Last annotation update) | |
| DE | Streptococcal histidine triad family protein. | | |
| GN | SAG1233. | | |
| OS | Streptococcus agalactiae (serotype V). | | |
| OC | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; | | |
| OC | Streptococcus. | | |
| NCBI_TaxID | =216466; | | |
| NC | | | |
| NR | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=2603 V/R / Serotype V; | | |
| RX | MEDLINE=2222988; PubMed=12200547; | | |
| RA | Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S., | | |
| RA | Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D., | | |
| RA | Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C., | | |
| RA | DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R., | | |
| RA | Radune D., Fedorova N.B., Scanlan R., Khouri H., Mulligan S., | | |
| RA | Carty H.A., Clive R.T., Van Aken S.E., Gill J.J., Scarbelli M., Mora M., | | |
| RA | Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D., | | |
| RA | Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G., | | |
| RA | Fraser C.M.; | | |
| RT | "Complete genome sequence and comparative genomic analysis of an | | |
| RT | emerging human pathogen, serotype V <i>Streptococcus agalactiae</i> ." | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002). | | |
| RM | EMBL; AE014248; AAN00111.1; . | | |
| DR | TIGR; SAG1233; . | | |
| DR | InterPro; IPR006270; Strep_his_triad. | | |
| DR | Pfam; PF04270; strep_his_triad; 6. | | |
| DR | TIGRFAMs; TIGR01363; strep_his_triad; 1. | | |
| KW | Complete proteome. | | |
| SQ | SEQUENCE 822 AA; 92400 MW; 8CCC8DF316727F98 CRC64; | | |
| Query Match 22.5%; Score 937; DB 16; Length 822; | | | |
| Best Local Similarity 29.5%; Pred. No. 3.6e-44; | | | |
| Matches 27; Conservative 98; Mismatches 235; Indels 318; Gaps 32 | | | |
| Qy | 1 | SYELGLYQATVKNENRVSVID---GKQATQKTENLTPEDEVSKREGINAEQIVIKITDQ | 57 |
| Db | 22 | SYQLGKHMGGLATKQNGIAYIDDSKGVKAPKT-NKTMQDISAEGISAEQIVVKITDQ | 80 |
| Qy | 58 | YVTSHGSHYHYNGKVPDYAIISPELLMKDNPYKLKDEIDVNEKVGKGVVVKDGYKYYVL | 117 |


```

Qy 756 GLRNNLTQIMDNNSIMAEKLLALLKGSNPSSVSKEK 794
Db 733 -----DKPKRQKQADK---SNENQOPSPSEKKEE 757

RESULT 13
Q8NZ82 PRELIMINARY; PRT; 823 AA.
AC Q8NZ82;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2003 (TREMBlrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein spyM18_2072.
GN SPYM18_2072.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AB010110; AAL98543.1; -.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 6.
DR TIGRFAMs; TIGR01363; strep_his_triad; 4.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 823 AA; 92585 MW; C79E1EB30CEBDF0C CRC64;

Query Match 22.7%; Score 946; DB 16; Length 823;
Best Local Similarity 29.9%; Pred. No. 1.1e-44;
Matches 261; Conservative 123; Mismatches 272; Indels 216; Gaps 29;

Qy 1 SYELGLQARTVKENRVSYID---GKQATQKTENLTNPDEVSKREGINAEQIVIKITDQG 57
Db 22 SYQLGKHHMGVTKDNQIAYIDDSKGAAPKT-NKTMDQISAEEGISAEQIVVKITDQG 80

Qy 58 YVTSHGHDHYHYNGKVPYDAIISBEELMKDPNYKLKDEDIVNEKGGYVIVKDGKYYVYL 117
Db 81 YVTSHGHDHYHYNGKVPYDAIISBEELMKDPNYKQSDVINEILDGYVIVKVGNGYYVYL 140

Qy 118 KDAHADNVRTKEINRQKQHSQH-REGG-----TPRNDGVALARSQGRVTTDDGY 169
Db 141 KPGSKRKNIRTKQIAEQVAKGTKEAKEKGLAQVAHLSKEEVAVNEAKRQGRVTTDDGY 200

Qy 170 IFNASDIIEDTGDAIYVPHGPHYVHPKNSLSASLAAAEFLS---GRGNLSRSTYRR 226
Db 201 IFSFTDIIDLDGAYLVPNGHYHYVHPKDLSPSELAAQYWSQKQGRG--ARPSDYRP 258

Qy 227 QNSNTSRKTNWVPSVS-----NPGTNTNTNSNTNSQASQSNIDSLK 272
Db 259 TPAPGRKKAP-IPDVTNPNRQHQPNDGGYHPAPPRPNDASQNKHQDEFKGTFKELLD 317

Qy 273 QLYKLPLSORHVESDGLVDFDPAQITSRARGVAVPHGPHYHYFIPYSQMSLEERARIIP 332
Db 318 QLHRLDLKYRVEEDGLFEPTQVKSNAFYVVPNGHYHYIIPRSQISPLEMELAD--- 374

Qy 333 LRYRSNHWVPDRPQSPQPTPEPSPGPQAPNLKIDNSLSVLSQVVRKVGEGVVPPEK 392
Db 375 -RYLAG-----QTEDNDSGDHDKPSDK-EVTHFFIGHRIKAYGKGLDGXPY 419

Qy 393 GIS-RYVFAKLPSETVKNLSKLSQBSVSGHTLTAKENVAPRQDFYDKAYNLLTEAH 451
Db 420 DTSDAYVFSK-----ESIHSV-----KSGVATAKGDHPHYIGFGELEQYE 460

```

RESULT 14

```

Q8E4U1 PRELIMINARY; PRT; 822 AA.
AC Q8E4U1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN GBS1306.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=2242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Meadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766850; CAD46965.1; -.
DR SagalList; gbs1306; -.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 6.
DR TIGRFAMs; TIGR01363; strep_his_triad; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 822 AA; 92393 MW; 2929A97C8AFCD78F CRC64;

```

Query Match 22.7%; Score 945; DB 16; Length 822;
 Best Local Similarity 30.7%; Pred. No. 1.3e-44;
 Matches 269; Conservative 117; Mismatches 285; Indels 226; Gaps 33;

```

Qy 1 SYELGLQARTVKENRVSYID---GKQATQKTENLTNPDEVSKREGINAEQIVIKITDQG 57
Db 22 SYQLGKHHMGVTKDNQIAYIDDSKGAAPKT-NKTMDQISAEEGISAEQIVVKITDQG 80

Qy 58 YVTSHGHDHYHYNGKVPYDAIISBEELMKDPNYKLKDEDIVNEKGGYVIVKDGKYYVYL 117
Db 81 YVTSHGHDHYHYNGKVPYDAIISBEELMKDPNYKQSDVINEILDGYVIVKVGNGYYVYL 140

```


| | | | |
|----|-----|--|-----|
| Qy | 301 | ARGVAVPHGDHYHFIPIYSQMSLEERARIILPIRYSRNHVVPDRPQSPQPTPEPSPG | 360 |
| Db | 302 | PNGVAIPHGHDYHFIPIYSKLSALEEKIARWVP | 333 |
| Qy | 361 | POPAPNLKIDSSSLVSQLVKRVKGVGVPEEKGISRVFAKDLPSFTVKNLKESKLSQES | 420 |
| Db | 334 | -----ISGTGSTVSTNAK-----PNEVVSSLSGLSSNPSS | 363 |
| Qy | 421 | VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGRRNSQFQALDKULLERLNDESTN | 480 |
| Db | 364 | -----LATSCKLSASDGYIENPK-DIVEETATAYIVRHG--DHPHYIPK-----SNQIG | 410 |
| Qy | 481 | KEKLVDDLLAFLAPIHPERLGRKPSQIETYTEDEVRIAQLADKYTTSDGYIFDEHDIISD | 540 |
| Db | 411 | OPTLPNNLSLATPSP-SLPINFGTSHEKHE-----EDGYGFDANRLIAE | 452 |
| Qy | 541 | EGDAYVTPHMGHSHWIKGSLDSKEKVAQAQYTKKGILPPSPDADYKANPTGDSAAAIY | 600 |
| Db | 453 | DESGFVMSHGDNHNYFPFKDLTTEQIKAAQKHLKE-----VKTSHNGLDLSLSSH | 503 |
| Qy | 601 | NR-----VKGEKRIPLVRLPYMV--EHTVEVKNGNLLII-PKHQVHNINIKFAW | 644 |
| Db | 502 | EQDYPSNAKEMKOLDKKIEBKIAIGMKQYGVKRESIVVNKEKNAILIYPHGDHHAADP--- | 558 |
| Qy | 645 | FDDHITKAPNGYLTLEDLFAIKYKVVVEHPDRPHSNDG-----WGNASEHVLGCKDHSEDPN | 700 |
| Db | 559 | IDEH---KPVGIG-----HSHSNVELFKPEGVAKCKGNKYVTGELTNVNNLLKNSTFNN | 611 |
| Qy | 701 | KNF 703 | |
| Db | 612 | QNF 614 | |

RESULT 11

Q8DQ07 PRELIMINARY; PRT; 1039 AA.

| | | | |
|----|---|--------------|---------------|
| ID | Q8DQ07 | PRELIMINARY; | PRT; 1039 AA. |
| AC | Q8DQ07; | | |
| DT | 01-MAR-2003 (TrEMBLrel. 23, Created) | | |
| DT | 01-MAR-2003 (TrEMBLrel. 23, Last sequence update) | | |
| DT | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) | | |
| DE | Pneumococcal histidine triad protein E. | | |
| GN | PHTE OC SPR0908. | | |
| OS | Streptococcus pneumoniae (strain ATCC BAA-255 / R6). | | |
| OC | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; | | |
| OC | Streptococcus. | | |
| OX | NCBI_TaxID=171101; | | |
| EN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=21429245; PubMed=11544234; | | |
| RA | Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S., | | |
| RA | DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C., | | |
| RA | Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E., | | |
| RA | LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P., | | |
| RA | McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I., | | |
| RA | Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P., | | |
| RA | Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G., | | |
| RA | Zook C.A., Bantz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L., | | |
| RA | Glass J.I.; | | |
| RT | "Genome of the bacterium Streptococcus pneumoniae strain R6.;" | | |
| RL | J. Bacteriol. 183:5709-5717(2001). | | |
| DR | EMBL; AE008464; AAK99712.1; -. | | |
| DR | PIR; D97985; D97985. | | |
| DR | InterPro; IPR006270; Strep_his_triad. | | |
| DR | Pfam; PF04270; strep_his_triad; 6. | | |
| DR | TIGRFAMs; TIGR01363; strep_his_triad; 3. | | |
| KW | Complete proteome. | | |
| SQ | SEQUENCE 1039 AA; 114625 MW; 05CC226D2028F551 CRC64; | | |

Query Match 29.8%; Score 1243; DB 16; Length 1039;

Best Local Similarity 41.2%; Pred. No. 3.7e-61;

Matches 298; Conservative 94; Mismatches 180; Indels 152; Gaps 20;

```
Db 141 AAHADNIRTKBEIKRQKQSHSHNHGGGS--NQAVVAARAQGRYTTDDGQYVFNASDIIED 198
QY 180 TGDYAVPHGDHYHYIPKNELSASELAARAEFLSGRGNLSNRTYRONSNDTNRNWPV 239
Db 199 TGDYAVPHGDHYHYIPKNELSASELAARAEFLSGRGNLSNRTYRONSNDTNRNWPV 247
QY 240 SVSNFGTNTNTSNNSTNSQASQSDNDISLLKQLYKLPQSQRHVESDGLVFPDPAQITSR 299
Db 248 NPAQPLSENHNLTVTPTVHQ--NOGENISLLRELKAKLPSEHVESDGLVFPDPAQITSR 306
QY 300 TARGVAVPHGDHYHYIPYQMSLEBRIARIIPLYRSNHWVPDSPEQSPQPTPEPSP 359
Db 307 TARGVAVPHGDHYHYIPYQMSLEBRIARIIPLYRSNHWVPDSPEQSPQPTPEPSP 366
QY 360 GPQAPNLK-IDSN---SSIVSOLVRKVGEGYVEEKGISRYVFAKDLSETVKNLESKL 415
Db 367 SQQAPNPQAPSNPIDKLVKAEVRKVGQYVEENGVSRYIPAKDLSEATAAGIDSKL 426
QY 416 SKQESVSHLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLERLN 475
Db 427 AKQESLSHKLGAKTDLPSDDREFYNKAYDILLARIHQDLLDNKGRQVDFEALDNLLERLK 486
QY 476 DESTNKEKLVDDLLAFIAPITHPERLGKNSOIEYTEDVRIAQLADKYTTSDGYTFDEH 535
Db 487 DVPSDKVLVDLLAFIAPIRHPERLGKNAQITTYTDDBIQVAKLAGKYTTEDGYTFDPR 546
QY 536 DIISDEGDYVTPHMGSHWIGKDSLDKEKVAARQAYTKEGILPPSPADVKANPTGDS 595
Db 547 DITSDEGDYVTPHMGSHWIGKDSLDKEKVAARQAYTKEGILPPSPADVKANPTGDS 606
QY 596 AAAIYNRVKGEKRIPLVRIPYVMEHTVEYKNGNLIIPHDXHYNIKFAMFDDHTYKAPNG 655
Db 607 AEAIYNRVKAAKVPDLRMPYNLYQTVVEKNGSLIIPHYDHYHNKFEWFDEGLYEAPKG 666
QY 656 YTELDLPAIKYVVEHDPDRPHSDNGWGNASHVLGK-----KDHSE----- 697
Db 667 YTELDLPAIKYVVEHDPDRPHSDNGWGNASHVLGK-----KDHSE----- 726
QY 698 -----DPKNFKADEPEVEETPAEPEVQVETKEVYEAQLKAEVLL 738
Db 727 ESEKENIAGLPSADNLKPSDTEETEERAEADTTDEAEIPOVENSVINAKIADAEALL 786
QY 739 AKVTDSSLKANATETLAGLNNLTQIMDNNSIMAEAEKLLALLKGSNPSV 790
Db 787 EKVTPDSIRQNAWETLTGLKSLILGTKNNTISAEDVSLALLKESQAPI 838

RESULT 9
QBDQ08
ID Q8DQ08 PRELIMINARY; PRT; 853 AA.
AC Q8DQ08;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Pnuemococcal histidine triad protein D.
GN PHTD OR SPR0907.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE 129245; PubMed=11544234;
RX MEDLINE=21429245;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burtett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAlharen S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
```

```
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008464; AAK99711.1; -.
DR PIR; C97985; C97985.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRPFAMs; TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 853 AA; 95225 MW; 9BF06A1EDEB90CF5 CRC64;

Query Match 65.6%; Score 2733; DB 16; Length 853;
Best Local Similarity 63.5%; Pred. No. 1.1e-144;
Matches 538; Conservative 92; Mismatches 145; Indels 72; Gaps 9;

QY 1 SYELGYYQARTV-KENNRVSYIDGKOATKNTLTPDEVSKRGINAEQIVIKITDQGYV 59
Db 21 SYELGHRQAGVKESNRVSYIDGQOQRAENLTPEVSKRGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDEIVNEVKGGYVVKDGYVYLKD 119
Db 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDEIVNEIKGGYVVKDGYVYLKD 140
QY 120 AAHADNVRKKEINROKQESHRREGGTPRNDCAVALARSQRYTTDDGYIFNASDIIED 179
Db 141 AAHADNIRTKBEIKRQKQSHSHNHNS---RADNAVAARAQGRYTTDDGYIFNASDIIED 197
QY 180 TGDYAVPHGDHYHYIPKNELSASELAARAEFLSGRGNLSNRTYRONSNDTNRNWPV 239
Db 198 TGDYAVPHGDHYHYIPKSDLSASELAARAAQAWNG-----KQSRPSSSSSHNA 246
QY 240 SVSNFGTNTNTSNNSTNSQASQSDNDISLLKQLYKLPQSQRHVESDGLVFPDPAQITSR 299
Db 247 NPAQPLSENHNLTVTPTVHQ--NOGENISLLRELKAKLPSEHVESDGLVFPDPAQITSR 305
QY 300 TARGVAVPHGDHYHYIPYQMSLEBRIARIIPLYRSNHWVPDSPEQSPQPTPEPSP 359
Db 306 TARGVAVPHGDHYHYIPYQMSLEBRIARIIPLYRSNHWVPDSPEQSPQPTPEPSP 365
QY 360 GPQAPNLK-IDSN---SSIVSOLVRKVGEGYVEEKGISRYVFAKDLSETVKNLESKL 415
Db 366 SQQAPNPQAPSNPIDKLVKAEVRKVGQYVEENGVSRYIPAKDLSEATAAGIDSKL 425
QY 416 SKQESVSHLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLERLN 475
Db 426 AKQESLSHKLGAKTDLPSDDREFYNKAYDILLARIHQDLLDNKGRQVDFEALDNLLERLK 485
QY 476 DESTNKEKLVDDLLAFIAPITHPERLGKNSOIEYTEDVRIAQLADKYTTSDGYTFDEH 535
Db 486 DVPSDKVLVDLLAFIAPIRHPERLGKNAQITTYTDDBIQVAKLAGKYTTEDGYTFDPR 545
QY 536 DIISDEGDYVTPHMGSHWIGKDSLDKEKVAARQAYTKEGILPPSPADVKANPTGDS 595
Db 546 DITSDEGDYVTPHMGSHWIGKDSLDSEAEARAAQAYAKEGLTPSTDHQSGNTEAKG 605
QY 596 AAAIYNRVKGEKRIPLVRIPYVMEHTVEYKNGNLIIPHDXHYNIKFAMFDDHTYKAPNG 655
Db 606 AEAIYNRVKAAKVPDLRMPYNLYQTVVEKNGSLIIPHYDHYHNKFEWFDEGLYEAPKG 665
QY 656 YTELDLPAIKYVVEHDPDRPHSDNGWGNASHVLGK-----DHSEDPNK----- 701
Db 666 YTELDLPAIKYVVEHDPDRPHSDNGWGNASHVLGK-----DHSEDPNK----- 725
QY 702 -----NFKAD-----EPEVEETPAEPEVQVE 723
Db 726 EEDKEHDEVSEPTHESEDEKENHVGILNPSADNLKPSDTEETEERAEADTTDEAEIPOVE 785
QY 724 TEKVBAQLKAEVLLAKVTDSSLKANATETLAGLNNLTQIMDNNSIMAEAEKLLALLK 783
Db 786 HSVINAKIAEAEALKEKVTDSIRQNAWETLTGLKSLILGTKNNTISAEDVSLALLK 845
QY 784 GSNPSSV 790
Db 846 ESQPTPI 852
```

```
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dornitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RL proteins (the Pht family) that are protective against sepsis.";
DR EMBL; AF318954; AAK06759.1; -
DR InterPro; IPR006270; Strep_his_triad; 5.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW Signal.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 819 819
SQ SEQUENCE 819 AA; 92108 MW; E602CFC16CC28A5F CRC64;

Query Match 66.6%; Score 2786.5; DB 2; Length 819;
Best Local Similarity 66.7%; Pred. No. 1.1e-147;
Matches 544; Conservative 84; Mismatches 140; Indels 47; Gaps 7;

QY 1 SYELGLYQA-RTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYV 59
DB 21 SYELGRVQAGDQKESNRVAIDGDAQKAENLTPEVSKREGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLDEIVNEVKGYVIKVDGKYVYLKD 119
DB 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKSDIIVNEIKGVIKVGKYYVYLKD 140
QY 120 AAHADNVRTKEEINRQKQHSQHREGTTPRDNGAVALARSQGRYTTDDGYIFNASDIIE 179
DB 141 AAHADNVRTKEEIRKQKQHSQHNS---RADNAVAARAAGRYTTDDGYIFNASDIIE 197
QY 180 TGDYIVPHGDHYHYIPKNSLSASELAAAEAFLSGRGNLSNRTYVRGNSDNTSRTNWP 239
DB 198 TGDYIVPHGDHYHYIPKNSLSASELAAAEAYWNG-----KQGRPSSSSSYNA 246
QY 240 SVSNPGTNTNTSNNTNSQASQSDIDSLKQYKLPQSRHVESDGLVDPQAITSR 299
DB 247 NPAQPRLSNHNLTVPPTYHQ-NQGENISLLRELYAKPLSERHVESDGLIFDPAITSR 305
QY 300 TARGVAVPHGDHYHFIYSQMSLEERLARIIPLYRSNHWVPSRPEQSPQPTPEPSP 359
DB 306 TARGVAVPHGNHYHFIPEYQMSLEKRIARIIPLYRSNHWVPSRPEEPSPQPTPEPSP 365
QY 360 GPQAPNLIKDSNSLSVQLVRKVGEGYVFEKGIISRYVFAKDLPSFTVKNLKSLKQE 419
DB 366 SPQAPSNPID--GKLVKAEVRKVGDGVPFENGVSRYIPAKDLSAETAAGIDSKLAKQE 423
QY 420 SVSHTLTAKENAVPRQDFVDKAYNLLTEAHKLFNKNRNSQFQALDKLLERLNDEST 479
DB 424 SLSHKLGTKTKDLPSSDREFYKNQYDILLARHQDLNKGQVDFEALDNLRLKDVSS 483
QY 480 NKEKLVDDLLAFAPITHPERLKPNSQIEVTEDEVRLAQADKYTTSDGYIFPEHDIIS 539
DB 484 DKVKLVDEDLIAFLAPIRHPERLKPNAQITVTDDEIQVAKLAGKYTAEDGDYIFDPDITS 543
QY 540 DEGDAYVTPHMGHSHWIKGDSLSDEKVAQAAYTKKGIILPPSPDADVKNAPTGDSSAAAI 599
DB 544 DEGDAYVTPHMTSHWIKKDSLSSEARAAQAAYAXEKLTPPSTDHQDSNGTEAKGEAI 603
QY 600 YNRVKGKRIPLVRLPVMVHTVVKXGNLIIPKHQVHNLIKFAWFDHTYKAPNGYTL 659
DB 604 YNRVKAACKVPLDRMPYNLQYTVVEVKGSLIIPHYDHYHNLIKFEWFDGLYAPKGYTL 663
QY 660 DLFATIKYVVEHPDRPHSDNGMGNASEHVLGKDHSDPDNKNFKADEE-----PVEET 713
DB 664 DLLATVIKYVVEHPNRPHSNGFNGASDHVORNKNGQADTNQTEKPEEKQTEKPEET 723

QY 714 -----PAEPEVPOVETEKVEAQLKEAEVLLAKVTTDSLSKANA 750
DB 724 PREEKPOSEKPEPKPTEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEE 783
QY 751 TETLAGLRNNLTQIMDNNSIMAEAEKLLALLKGS 785
DB 784 KETLTGLKNLLFTGTDNNNTIMAEAEKLLALLKES 818

RESULT 8
Q9ANY2 PRELIMINARY; PRT; 839 AA.
ID Q9ANY2;
AC Q9ANY2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 24, Last annotation update)
DE Pneumococcal histidine triad protein D precursor (Hypothetical protein
DE SP1003) (Fragment).
GN PHTD OR SP1003.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dornitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RL proteins (the Pht family) that are protective against sepsis.";
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heideberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang P., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AF318955; AAK06760.1; -
DR EMBL; AE007403; AAK75120.1; -
DR PIR; G95115; G95115.
DR TIGR; SP1003; -
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW Signal; Hypothetical protein; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;

Query Match 66.6%; Score 2772; DB 16; Length 839;
Best Local Similarity 65.0%; Pred. No. 7e-147;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLYQARTV-KENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYV 59
DB 21 SYELGRHQAGQVKKESNRVSYIDGDAQKAENLTPEVSKREGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLDEIVNEVKGYVIKVDGKYVYLKD 119
DB 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKSDIIVNEIKGVIKVGKYYVYLKD 140
QY 120 AAHADNVRTKEEINRQKQHSQHREGTTPRDNGAVALARSQGRYTTDDGYIFNASDIIE 179
```

QY 301 ARGVAVPHGDHVFIPYSOMSELEERARIIPLYRSNHWVPDSRPEQSPQTPPSPG 360
 Db 332 ANGVAVPHGDHVFIPYSQSPLEEKARIIPLYRSNHWVPDSRPEQSPQTPPSPS 391
 QY 361 POPAPNLK-IDSN---SSLSQLVRKVGEGYVFEKGISRYVFAKOLPSTVKNLESKLS 416
 Db 392 POPAPNPQAPSNPIDKLVKAVRKVGEGYVFEENGVPRIIPAKDLSAETAAGIDSKLA 451
 QY 417 KQSVSHTLTAKENVAPRDQBYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLND 476
 Db 452 KQESLHKLGAKTDLPSDDREFYNKAYDILLARIHQDLLDNKGRQVDFEALDNLLEKLD 511
 QY 477 ESTNKEKLVDDLLAFAPITHPERLGNPSQIETVDEVRIAQLADKYTTSDGYIFDEHD 536
 Db 512 VSDKVKLVDDLLAFAPITHPERLGNPNQIYTTDDEIQVAKLAGKYTTEDGYIFDPRD 571
 QY 537 IISDEGDVYTPHMGSHWIGKDSLSDKEKVAQAAYTEKGIILPPSPADAVKANPTGDSA 596
 Db 572 ITSDEGDVYTPHMTSHWIKKDSLSEAEARAAQAYAKEKGLTPPSTDHQDSGNTAKGA 631
 QY 597 AAIYNRVKEKRIPIVRLPVMVHEHTVEVKNGLIIPHKDHYHNIKFAWPDHHTYKAPNGY 656
 Db 632 EAIYNRVKAACKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNIKFEWFDGLYEAPKGY 691
 QY 657 TLEDLPATIKYVVEHPDERPHSDNGMGNASEHVLGKKHSDDPKNKFKADEE-----PV 710
 Db 692 SLEDLATKYVVEHPNERPHSDNGMGNASHVQRNKGQADNTQTEKNEEYPQTEKPE 751
 QY 711 EET-----PAPPEVPQVETEKVAQLKEAEVLLAKVT 742
 Db 752 EETPREKQSEKPEPKPTEPEERSPESEEPQVETEKVKEKLEAEADLLGKIQ 811
 QY 743 DSSLKANAVETLAGLNLTLOIMDNNSIMAEKLLALLKGS 785
 Db 812 NPIIKSNKATETLGNKLLFGTDNNIMAEKLLALLKES 854
 RESULT 6
 Q97QM9 PRELIMINARY; PRT; 819 AA.
 AC Q97QM9;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Conserved domain protein.
 GN SP1174.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]_TaxID=1313;
 RC SEQUENCE FROM N.A.
 RX STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapfle E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollinghead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae.";
 RL Science 293:498-506 (2001).
 DR EMBL; AB007418; AAK75283.1; -.
 DR PIR; B95136; B95136.
 DR TIGR; SP1174; -.
 DR InterPro; IPR006270; Strep_his_triad.
 DR Pfam; PF04270; strep_his_triad; 5.
 DR TIGRFAMS; TIGR01363; strep_his_triad; 2.
 KW Complete proteome.

SQ SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;
 Query Match 67.2%; Score 2797.5; DB 16; Length 819;
 Best Local Similarity 67.0%; Pred. No. 2.5e-148;
 Matches 546; Conservative 84; Mismatches 138; Indels 47; Gaps 7;
 QY 1 SYELGLYQA-RTVKENNRVSYIDGKQATOKTENLTLPDEVSKREGINAEQIVIKITDQYV 59
 Db 21 SYELGRYQAGQDKKESNRVAYIDGQAGQKAENLTDEVSKREGINAEQIVIKITDQYV 80
 QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNPKYKDEDIVNEVKGVIKVDGKYVYVLKD 119
 Db 81 TSHGDHYHYNGKVPYDAIISELLMKDPNPKYKDSDIVNEIKGGYVIVKNGYVYVLKD 140
 QY 120 AAHADNVRTKEINRQKQHSQHSQREGTPRNDCAVALARSQGYTTDDGYIFNASDIIE 179
 Db 141 AAHADNIRIKETIKRQKQERSHNHS-----RADNAVAARAAQGYTTDDGYIFNASDIIE 197
 QY 180 TGDAYIVPHGDHYHYIPKNELSASELAAAFISGRGNLSNRTYRQNSDNTSRTNWVP 239
 Db 198 TGDAYIVPHGDHYHYIPKNELSASELAAAEAYWNG-----KQGSRPSSSSSYNA 246
 QY 240 SVSNPGTNTNTNSNTNSQASQSDIDSLKQLYKLPLSQRHVESDGLVFPDPAQITSR 299
 Db 247 NPAQRLSENHNLTVPTVHQ-NQGENISLLRELYAKPLSERHVESDGLVFPDPAQITSR 305
 QY 300 TAGVAVPHGDHYHYFIPYSOMSELEERARIIPLYRSNHWVPDSRPEQSPQTPPSP 359
 Db 306 TAGVAVPHGNHVFIPYEQMSELEERARIIPLYRSNHWVPDSRPEQSPQTPPSP 365
 QY 360 GPQAPNLKIDSNSLSVLVRKVGEGYVFEKGISRYVFAKOLPSTVKNLESKLSKQE 419
 Db 366 SPQAPFNPID--EKLVAEAVRKVGEGYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQE 423
 QY 420 SVSHHTTAKENVAPRDQBYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLND 479
 Db 424 SLSHKLGTKTKTDLPSDDREFYNKAYDILLARIHQDLLDNKGRQVDFEALDNLLEKLDVSS 483
 QY 480 NKEKLVDDLLAFAPITHPERLGNPSQIETVDEVRIAQLADKYTTSDGYIFDEHDII 539
 Db 484 DKVKLVEDLLAFAPITHPERLGNPNQIYTTDDEIQVAKLAGKYTTEDGYIFDPRDITS 543
 QY 540 DEGDVYTPHMGSHWIGKDSLSDKEKVAQAAYTEKGIILPPSPADAVKANPTGDSAAA 599
 Db 544 DEGDVYTPHMTSHWIKKDSLSEAEARAAQAYAKEKGLTPPSTDHQDSGNTAKGAEL 603
 QY 600 YNRVKEKRIPIVRLPVMVHEHTVEVKNGLIIPHKDHYHNIKFAWPDHHTYKAPNGY 659
 Db 604 YNRVKAACKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNIKFEWFDGLYEAPKGY 663
 QY 660 DLFATIKYVVEHPDERPHSDNGMGNASEHVLGKKHSDDPKNKFKADEE-----PVEET 713
 Db 664 DLLATKYVVEHPNERPHSDNGMGNASHVQRNKGQADNTQTEKNEEYPQTEKPEET 723
 QY 714 -----PAPPEVPQVETEKVAQLKEAEVLLAKVTSSSLKANA 750
 Db 724 PREKQSEKPEPKPTEPEERSPESEEPQVETEKVKEKLEAEADLLGKIQDPIIKSNA 783
 QY 751 TETLAGLNLTLOIMDNNSIMAEKLLALLKGS 785
 Db 784 KETLTGLKNLLFGTDNNIMAEKLLALLKES 818
 RESULT 7
 Q9ANY3 PRELIMINARY; PRT; 819 AA.
 ID Q9ANY3;
 AC Q9ANY3;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Pneumococcal histidine triad protein B precursor (Fragment).
 GN PHTB.
 OS Streptococcus pneumoniae.

121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
Db 141 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 200
QY 181 GDAYIVPHGDHGHYIIPKNELASSELAAAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240
Db 201 GDAYIVPHGDHGHYIIPKNELASSELAAAFISGRGNLSNRTYRQNSDNTSRTNWVPS 260
QY 241 VSNPGTNTNTSNNSTNSQASNSDIDSLKQYKLPISQRHVSDGLVDFDAQTTSRT 300
Db 261 VSNPGTNTNTSNNSTNSQASNSDIDSLKQYKLPISQRHVSDGLVDFDAQTTSRT 320
QY 301 ARGVAVPHGDHGHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 360
Db 321 ARGVAVPHGDHGHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 380
QY 361 POPAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKOLPSETVKNLESKLSKQES 420
Db 381 POPAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKOLPSETVKNLESKLSKQES 440
QY 421 VSHLTAKKENAVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db 441 VSHLTAKKENAVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
QY 481 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540
Db 501 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 560
QY 541 EGDAYVTPHMGSHHWIGKDSLSDEKVAQAAYTKKGIILPPSPDADVKNPTGDSAAAY 600
Db 561 EGDAYVTPHMGSHHWIGKDSLSDEKVAQAAYTKKGIILPPSPDADVKNPTGDSAAAY 620
QY 601 NRKVGKRIPLVRLPYMVEHTVEVKNGLIIPKHQHYHNIKFAWFDHHTYKAPNGYTLED 660
Db 621 NRKVGKRIPLVRLPYMVEHTVEVKNGLIIPKHQHYHNIKFAWFDHHTYKAPNGYTLED 680
QY 661 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNPKADEEPEEPTABPEVP 720
Db 681 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNPKADEEPEEPTABPEVP 740
QY 721 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 780
Db 741 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 800
QY 781 LKGSNPSSVSKEKIN 796
Db 801 LKGSNPSSVSKEKIN 816

RESULT 3
Q8DPQ2 PRELIMINARY; PRT; 828 AA.
ID Q8DPQ2;
AC Q8DPQ2;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Pneumococcal histidine triad protein A.
GN PHTA OR SPRL061.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.B. Jr., Arnold J., Blaszcak L.C., Bargett S.,
DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,

Glass J.I.;
"Genome of the bacterium Streptococcus pneumoniae strain R6.";
J. Bacteriol. 183:5709-5717(2001).
EMBL; AB008479; AK99865.1; -.
DR PIR; E98004; E98004.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 828 AA; 93015 MW; 12CCCF407B550CLD CRC64;
Query Match 99.9%; Score 4159; DB 16; Length 828;
Best Local Similarity 99.7%; Pred. No. 1.6e-224;
Matches 794; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 SYELGLYQARTVKNRNVSYIDGKQATQKTENITPDEVSKREGINAEQIVIKITDQGYVT 60
Db 33 SYELGLYQARTVKNRNVSYIDGKQATQKTENITPDEVSKREGINAEQIVIKITDQGYVT 92
QY 61 SHGDHYHYNGKVPYDAIISSELLMKDPNPKLDEDIVNEVGKGYIVKVDGKYYVYLKDA 120
Db 93 SHGDHYHYNGKVPYDAIISSELLMKDPNPKLDEDIVNEVGKGYIVKVDGKYYVYLKDA 152
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
Db 153 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 212
QY 181 GDAYIVPHGDHGHYIIPKNELASSELAAAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240
Db 213 GDAYIVPHGDHGHYIIPKNELASSELAAAFISGRGNLSNRTYRQNSDNTSRTNWVPS 272
QY 241 VSNPGTNTNTSNNSTNSQASNSDIDSLKQYKLPISQRHVSDGLVDFDAQTTSRT 300
Db 273 VSNPGTNTNTSNNSTNSQASNSDIDSLKQYKLPISQRHVSDGLVDFDAQTTSRT 332
QY 301 ARGVAVPHGDHGHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 360
Db 333 ARGVAVPHGDHGHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 392
QY 361 POPAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKOLPSETVKNLESKLSKQES 420
Db 393 POPAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKOLPSETVKNLESKLSKQES 452
QY 421 VSHLTAKKENAVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db 453 VSHLTAKKENAVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 512
QY 481 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540
Db 513 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 572
QY 541 EGDAYVTPHMGSHHWIGKDSLSDEKVAQAAYTKKGIILPPSPDADVKNPTGDSAAAY 600
Db 573 EGDAYVTPHMGSHHWIGKDSLSDEKVAQAAYTKKGIILPPSPDADVKNPTGDSAAAY 632
QY 601 NRKVGKRIPLVRLPYMVEHTVEVKNGLIIPKHQHYHNIKFAWFDHHTYKAPNGYTLED 660
Db 633 NRKVGKRIPLVRLPYMVEHTVEVKNGLIIPKHQHYHNIKFAWFDHHTYKAPNGYTLED 692
QY 661 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNPKADEEPEEPTABPEVP 720
Db 693 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNPKADEEPEEPTABPEVP 752
QY 721 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 780
Db 753 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 812
QY 781 LKGSNPSSVSKEKIN 796
Db 813 LKGSNPSSVSKEKIN 828

RESULT 4

```

90 155 3.7 869 4 Q9NYF8 Q9nyf8 homo sapien
ALIGNMENTS
RESULT 1
Q97QM8 PRELIMINARY; PRT; 802 AA.
AC Q97QM8;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Conserved domain protein.
GN SP1175.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,
RA Holtzapple E., Khouiri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AB007418; AAKV5284.1; -.
DR PIR; C95136; C95136.
DR TIGR; SP1175; -.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 802 AA; 90080 MW; 4F5CB8364EEA1833 CRC64;

Query Match 100.0%; Score 4163; DB 16; Length 802;
Best Local Similarity 99.9%; Pred. No. 9e-225;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 7 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 66
Qy 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIVNEKGGYVVKDGGYVYLKDA 120
Db 67 SHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIVNEKGGYVVKDGGYVYLKDA 126
Qy 121 AHADNVRTKEINRQKQHSQHRREGTTPNDGVALARSQRYTTDDGYIENASDIIDET 180
Db 127 AHADNVRTKEINRQKQHSQHRREGTTPNDGVALARSQRYTTDDGYIENASDIIDET 186
Qy 181 GDAYIVPHGDHYHYIPKNELASASLAAAEAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Db 187 GDAYIVPHGDHYHYIPKNELASASLAAAEAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 246
Qy 241 VSNPCTTNTNTSNNTNTSQASQNDIDSLKQLYKLPLSORHVESDGLVDPDAQITSR 300
Db 247 VSNPCTTNTNTSNNTNTSQASQNDIDSLKQLYKLPLSORHVESDGLVDPDAQITSR 306
Qy 301 ARGVAVPHGDHYHYIPYSOMSELEERIIARIIPLYRSNHNWVPDSRPEQSPQPTPEPS 360
Db 307 ARGVAVPHGDHYHYIPYSOMSELEERIIARIIPLYRSNHNWVPDSRPEQSPQPTPEPS 366
Qy 361 PQAPNLKIDSNSLVSQLVRKVGEGYVFEKGISRYVFAKDLPSETVKNLESKLSKQES 420

```

```

Db 367 PQAPNLKIDSNSLVSQLVRKVGEGYVFEKGISRYVFAKDLPSETVKNLESKLSKQES 426
Qy 421 VSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db 427 VSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLERLNDESTN 486
Qy 481 KEKLVDLLAFLAPITTHPERLGKPNKSIQIEYTEDEVRAQLADKYTTSDGYIFDEHDIISD 540
Db 487 KEKLVDLLAFLAPITTHPERLGKPNKSIQIEYTEDEVRAQLADKYTTSDGYIFDEHDIISD 546
Qy 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQAYTKEGILPPSPDADVKANPTGSAAYI 600
Db 547 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQAYTKEGILPPSPDADVKANPTGSAAYI 606
Qy 601 NRKGEKRIPLVRLPYMVEHTVEVKGNGLLIIPHKDHYHNIKFAWFFDDHTYKAPNGYTTLED 660
Db 607 NRKGEKRIPLVRLPYMVEHTVEVKGNGLLIIPHKDHYHNIKFAWFFDDHTYKAPNGYTTLED 666
Qy 661 LFATIKYVVEHPDERPHSDGNGNASBHVLGKDHSDPNKFNKADBPVEETPAEPEVP 720
Db 667 LFATIKYVVEHPDERPHSDGNGNASBHVLGKDHSDPNKFNKADBPVEETPAEPEVP 726
Qy 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETTLAQLRNNLTLOIMDNNSIMAEAKLLA 780
Db 727 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETTLAQLRNNLTLOIMDNNSIMAEAKLLA 786
Qy 781 LLKGSNPFSSVSKKIN 796
Db 787 LLKGSNPFSSVSKKIN 802

RESULT 2
Q9AHT9 PRELIMINARY; PRT; 816 AA.
AC Q9AHT9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Pneumococcal histidine triad A protein.
GN PHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsach C.,
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
RA Langermann S., Johnson S., Koenig S.;
RT "Use of a Whole Genome Approach To Identify Vaccine Molecules
RT Affording Protection against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1593-1598(2001).
DR EMBL; AF291695; AAK19155.1; -.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
SQ SEQUENCE 816 AA; 91519 MW; 5359126A611D27ED CRC64;

Query Match 100.0%; Score 4163; DB 2; Length 816;
Best Local Similarity 100.0%; Pred. No. 9.3e-225;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 21 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 80
Qy 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIVNEKGGYVVKDGGYVYLKDA 120
Db 81 SHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIVNEKGGYVVKDGGYVYLKDA 140

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 05:49:44 ; Search time 77 Seconds

(without alignments)
3261.721 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYELGLYQARTYKENRVSY.....KLALLKGSNPSSVSKKIN 796

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriapi.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 4163 | 100.0 | 802 | 16 | Q97QM8 streptococc |
| 2 | 4163 | 100.0 | 816 | 2 | Q9AHT9 streptococc |
| 3 | 4159 | 99.9 | 828 | 16 | Q8DPQ2 streptococc |
| 4 | 3236 | 77.7 | 844 | 2 | Q9AG74 streptococc |
| 5 | 3224 | 77.4 | 855 | 16 | Q8CWR4 streptococc |
| 6 | 2797.5 | 67.2 | 819 | 16 | Q97QM9 streptococc |
| 7 | 2786.5 | 66.9 | 819 | 2 | Q9ANY3 streptococc |
| 8 | 2772 | 66.6 | 839 | 16 | Q9ANY2 streptococc |
| 9 | 2733 | 65.6 | 853 | 16 | Q8DQ08 streptococc |
| 10 | 1246 | 29.9 | 1039 | 16 | Q9ANY1 streptococc |
| 11 | 1243 | 29.8 | 1039 | 16 | Q8DQ07 streptococc |
| 12 | 949 | 22.8 | 825 | 16 | Q99XV4 streptococc |
| 13 | 946 | 22.7 | 823 | 16 | Q8NZ82 streptococc |
| 14 | 945 | 22.7 | 822 | 16 | Q8E4U1 streptococc |
| 15 | 944 | 22.7 | 825 | 2 | Q93GT5 streptococc |
| 16 | 940 | 22.6 | 823 | 16 | Q8K5Q1 streptococc |

| | | | | | | |
|--------------------|----|-------|------|-------|----|--------|
| Q8dz81 streptococc | 17 | 937 | 22.5 | 822 | 16 | Q8DZ81 |
| Q9zbg7 streptococc | 18 | 934 | 22.4 | 822 | 2 | Q9ZBG7 |
| Q877Y2 streptococc | 19 | 922 | 22.1 | 794 | 16 | Q877Y2 |
| Q8e338 streptococc | 20 | 916.5 | 22.0 | 481 | 16 | Q8E338 |
| Q9ae1 streptococc | 21 | 679.5 | 16.3 | 289 | 2 | Q9AE21 |
| Q8esr2 streptococc | 22 | 305 | 7.3 | 877 | 16 | Q8ESR2 |
| Q8e029 streptococc | 23 | 302 | 7.3 | 877 | 16 | Q8E029 |
| Q8dq06 streptococc | 24 | 243 | 5.8 | 182 | 16 | Q8DQ06 |
| Q99z76 streptococc | 25 | 223 | 5.4 | 792 | 16 | Q99Z76 |
| Q8k714 streptococc | 26 | 213 | 5.1 | 792 | 16 | Q8K714 |
| Q963t1 plasmodium | 27 | 197.5 | 4.7 | 1078 | 5 | Q963T1 |
| Q77033 dictyosteli | 28 | 192.5 | 4.6 | 1390 | 5 | Q77033 |
| Q8p0g5 streptococc | 29 | 192 | 4.6 | 792 | 16 | Q8P0G5 |
| Q8ict8 plasmodium | 30 | 181.5 | 4.4 | 2752 | 5 | Q8ICT8 |
| Q15754 dictyosteli | 31 | 178 | 4.3 | 565 | 5 | Q15754 |
| Q8i156 plasmodium | 32 | 178 | 4.3 | 1233 | 5 | Q8I156 |
| Q8ib94 plasmodium | 33 | 177 | 4.2 | 8591 | 5 | Q8IB94 |
| Q25860 plasmodium | 34 | 176.5 | 4.2 | 1271 | 5 | Q25860 |
| Q9gtx2 plasmodium | 35 | 175.5 | 4.2 | 1236 | 5 | Q9GTX2 |
| Q8ihn3 plasmodium | 36 | 175.5 | 4.2 | 5507 | 5 | Q8IHN3 |
| Q8icv5 plasmodium | 37 | 173 | 4.2 | 1063 | 5 | Q8ICV5 |
| Q9u459 plasmodium | 38 | 172.5 | 4.1 | 5458 | 5 | Q9U459 |
| Q9lci9 staphylococ | 39 | 172 | 4.1 | 1795 | 16 | Q9LCJ9 |
| Q9lch2 staphylococ | 40 | 172 | 4.1 | 2478 | 2 | Q9LCH2 |
| Q99gr6 staphylococ | 41 | 172 | 4.1 | 2481 | 16 | Q99GR6 |
| Q8iea3 plasmodium | 42 | 171.5 | 4.1 | 3452 | 5 | Q8IEA3 |
| Q9rl69 staphylococ | 43 | 170 | 4.1 | 2478 | 2 | Q9RL69 |
| Q82345 arabidopsis | 44 | 168.5 | 4.0 | 1043 | 10 | Q82345 |
| Q8i413 plasmodium | 45 | 168.5 | 4.0 | 1708 | 5 | Q8I413 |
| Q7x0p4 lactobacill | 46 | 167.5 | 4.0 | 4427 | 2 | Q7X0P4 |
| Q8ieb6 plasmodium | 47 | 166.5 | 4.0 | 2651 | 5 | Q8IEB6 |
| Q8i223 plasmodium | 48 | 166 | 4.0 | 1465 | 5 | Q8I223 |
| Q97255 plasmodium | 49 | 166 | 4.0 | 1826 | 5 | Q97255 |
| Q9nfa3 drosophila | 50 | 166 | 4.0 | 16215 | 5 | Q9NFA3 |
| Q9i7u4 drosophila | 51 | 166 | 4.0 | 18074 | 5 | Q9I7U4 |
| Q8i3a8 plasmodium | 52 | 165.5 | 4.0 | 1373 | 5 | Q8I3A8 |
| Q8i121 plasmodium | 53 | 165.5 | 4.0 | 2227 | 5 | Q8I121 |
| Q8i2v4 plasmodium | 54 | 165.5 | 4.0 | 3381 | 5 | Q8I2V4 |
| Q99td3 staphylococ | 55 | 165 | 4.0 | 891 | 16 | Q99TD3 |
| Q8ze8 lactobacill | 56 | 165 | 4.0 | 1231 | 16 | Q8ZE8 |
| Q8i1n9 plasmodium | 57 | 165 | 4.0 | 3370 | 5 | Q8I1N9 |
| Q93lp4 staphylococ | 58 | 164 | 3.9 | 891 | 16 | Q93LP4 |
| Q8idp7 plasmodium | 59 | 162.5 | 3.9 | 1488 | 5 | Q8IDP7 |
| Q9u4x0 plasmodium | 60 | 162.5 | 3.9 | 2647 | 5 | Q9U4X0 |
| Q8nw39 staphylococ | 61 | 162 | 3.9 | 895 | 16 | Q8NW39 |
| Q7z3u2 homo sapien | 62 | 162 | 3.9 | 1118 | 4 | Q7Z3U2 |
| Q8i1d3 plasmodium | 63 | 162 | 3.9 | 2467 | 5 | Q8I1D3 |
| Q9fnd5 arabidopsis | 64 | 162 | 3.9 | 2910 | 10 | Q9FND5 |
| Q8i145 plasmodium | 65 | 162 | 3.9 | 3504 | 5 | Q8I145 |
| Q8i1z2 plasmodium | 66 | 162 | 3.9 | 4405 | 5 | Q8I1Z2 |
| Q9vpl9 drosophila | 67 | 162 | 3.9 | 5322 | 5 | Q9VPL9 |
| Q8i659 plasmodium | 68 | 161.5 | 3.9 | 1383 | 5 | Q8I659 |
| Q8ibl1 plasmodium | 69 | 160.5 | 3.9 | 1003 | 5 | Q8IBL1 |
| Q8i5h0 plasmodium | 70 | 160.5 | 3.9 | 1461 | 5 | Q8I5H0 |
| Q8cnu9 staphylococ | 71 | 160 | 3.8 | 3692 | 16 | Q8CNU9 |
| Q8x1l1 clostridium | 72 | 159.5 | 3.8 | 1129 | 16 | Q8X1L1 |
| Q8cp76 staphylococ | 73 | 159 | 3.8 | 9439 | 16 | Q8CP76 |
| Q83k2 enterococcu | 74 | 158.5 | 3.8 | 1499 | 16 | Q83K2 |
| Q77328 plasmodium | 75 | 158 | 3.8 | 1650 | 5 | Q77328 |
| Q8ied3 plasmodium | 76 | 158 | 3.8 | 1840 | 5 | Q8IED3 |
| Q8ic27 plasmodium | 77 | 157 | 3.8 | 1859 | 5 | Q8IC27 |
| Q8i1u7 plasmodium | 78 | 156.5 | 3.8 | 1139 | 5 | Q8I1U7 |
| Q8h6x1 arabidopsis | 79 | 156.5 | 3.8 | 1157 | 10 | Q8H6X1 |
| Q8ime5 drosophila | 80 | 156.5 | 3.8 | 1185 | 5 | Q8IME5 |
| Q9lza8 arabidopsis | 81 | 156.5 | 3.8 | 1495 | 10 | Q9LZA8 |
| Q9s1g8 arabidopsis | 82 | 156 | 3.7 | 519 | 10 | Q9S1G8 |
| Q8idj6 plasmodium | 83 | 156 | 3.7 | 1692 | 5 | Q8IDJ6 |
| Q7za38 ashbya goss | 84 | 156 | 3.7 | 3392 | 3 | Q7ZA38 |
| Q9p619 schizosacch | 85 | 155.5 | 3.7 | 1038 | 3 | Q9P619 |
| Q9veu8 drosophila | 86 | 155.5 | 3.7 | 1298 | 5 | Q9VEU8 |
| Q8ccb3 mus musculu | 87 | 155.5 | 3.7 | 1413 | 11 | Q8CCB3 |
| Q8i350 plasmodium | 88 | 155.5 | 3.7 | 3069 | 5 | Q8I350 |
| Q8ewu6 homo sapien | 89 | 155 | 3.7 | 689 | 4 | Q8EWU6 |

"The cloning of a receptor-type protein tyrosine phosphatase expressed in the central nervous system.";

J. Biol. Chem. 268:10573-10581(1993).
 [3]
 RT TISSUE=Liver;
 RN MEDLINE=91006018; PubMed=2170109;
 RX Krueger N.X., Streuli M., Saito H.;
 RA "Structural diversity and evolution of human receptor-like protein
 RT tyrosine phosphatases.";
 RL EMBO J. 9:3241-3252(1990).
 [4]
 RP SEQUENCE OF 1749-1990 AND 2047-2280 FROM N.A. (ISOFORM LONG).
 RC TISSUE=Brain stem; PubMed=2169617;
 RX MEDLINE=90384936; PubMed=2169617;
 RA Kaplan R., Morse B., Huebner K., Croce C., Howk R., Ravera M.,
 RA Ricca G., Jaye M., Schlössinger J.;
 RT "Cloning of three human tyrosine phosphatases reveals a multigene
 RT family of receptor-linked protein-tyrosine-phosphatases expressed in
 RT brain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7000-7004(1990).
 CC -!- FUNCTION: May be involved in the regulation of specific
 CC developmental processes in the CNS.
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
 CC tyrosine + phosphate.
 CC -!- SUBUNIT: The carbonic-anhydrase like domain binds to contactin (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Names=Long;
 CC IsoId=P23471-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P23471-2; Sequence=VSP_005151;
 CC -!- TISSUE SPECIFICITY: Specifically expressed in the central nervous
 CC system, where it is localized in the Purkinje cell layer of the
 CC cerebellum, the dentate gyrus, and the subependymal layer of the
 CC anterior horn of the lateral ventricle. Developmentally regulated
 CC in the brain.
 CC -!- SIMILARITY: Contains 1 eukaryotic-type carbonic anhydrase domain.
 CC -!- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
 CC -!- CAUTION: Called RPPase beta in Ref.2 and Ref.4.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M93426; AAA60225.1; -;
 CC EMBL; X54135; CAA38070.1; -;
 CC PIR; A46151; A46151.
 CC HSSP; P18052; 1YFO.
 CC Genew; HGNC:9685; PTPRZ1.
 CC MIM; 176891; -;
 CC GO; GO:0005887; C:integral to plasma membrane; TAS.
 CC GO; GO:0005001; F:transmembrane receptor protein tyrosine pho. . . ; TAS.
 CC GO; GO:0007417; P:central nervous system development; TAS.
 CC GO; GO:0006470; P:protein amino acid dephosphorylation; TAS.
 CC InterPro; IPR001148; Euk_Coahnd.
 CC InterPro; IPR008957; FN_III-like.
 CC InterPro; IPR003961; FN_III.
 CC InterPro; IPR003387; TYR_phosphatase.
 CC InterPro; IPR000242; Tyr_PP.
 CC Pfam; PF00194; carb_anhydrase; 1.
 CC Pfam; PF00041; fn3_1.
 CC Pfam; PF00102; Y_phosphatase; 2.
 CC PRINTS; PR00700; PRTYPHPHASE.
 CC ProDom; PD000865; Euk_Coahnd; 1.
 CC SMART; SM00060; FN3; 1.

DR SMART; SM00194; PTPC; 2.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
 DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
 KW Signal; glycoprotein; Transmembrane; Hydrolase; Repeat;
 KW Alternative splicing.
 FT SIGNAL 1 24 BY SIMILARITY.
 FT CHAIN 25 2314 RECEPTOR-TYPE PROTEIN-TYROSINE
 FT PHOSPHATASE ZETA.
 FT DOMAIN 25 1635 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1636 1661 POTENTIAL.
 FT DOMAIN 1662 2314 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 34 302 CARBONIC-ANHYDRASE LIKE.
 FT DOMAIN 312 406 FIBRONECTIN TYPE-III.
 FT DOMAIN 1744 1997 PROTEIN-TYROSINE PHOSPHATASE 1.
 FT DOMAIN 1998 2314 PROTEIN-TYROSINE PHOSPHATASE 2.
 FT ACT_SITE 1932 1932 PHOSPHOCYSTEINE INTERMEDIATE (BY
 FT SIMILARITY).
 FT SITE 2222 2222 ANCESTRAL ACTIVE SITE.
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 587 587 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)
 FT (POTENTIAL).
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 637 637 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)
 FT (POTENTIAL).
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 997 997 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)
 FT (POTENTIAL).
 FT CARBOHYD 1017 1017 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1082 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1122 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1456 1456 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1548 1548 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)
 FT (POTENTIAL).
 FT CARBOHYD 1550 1550 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)
 FT (POTENTIAL).
 FT CARBOHYD 1561 1561 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1617 1617 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 755 1614 Missing (in isoform Short).
 FT /FTId=VSP_005151.
 FT CONFLICT 1722 1728 MISSING (IN REF. 2).
 FT SEQUENCE 2314 AA; 254528 MW; 77DBDEF4A0F5FB42 CRC64;
 Query Match 3.4%; Score 141.5; DB 1; Length 2314;
 Best Local Similarity 19.2%; Pred. No. 13;
 Matches 169; Conservative 113; Mismatches 267; Indels 331; Gaps 43;
 QY 51 IKITDQGYVTHSGDHYHYNGKVPYDAIISEELMKDPNPKLXDEIVNEVKGYVIKVD 110
 Db 954 VGVTYQGSLSFGSPSHI-----PIPKSSLITPTASLLQPTHALSG-----D 993
 QY 111 GKYYVVLKDAHADNVRTKEINRQKHSQHREGGTGRNDGAVAL-----ARSGRYT 164
 Db 994 GEW-----SGASDS-----EFLLPDTGLTALNISPSVSAFTYT 1030
 QY 165 T-----DDGYIFNADSIIEDTDGAYIVPHGDHVVH-----IPK-----NELSASELAAAE 209
 Db 1031 TSVFGDDNKALSKEIILYGNTELOIPSNEMVYPSSEIWMYDYNVKLASLETSTV 1090
 QY 210 AFLSGRG-----NLNSRTYR-RQNSDNTS-----RTNWVP 239
 Db 1091 SISSTKGMFPFSLAHTTTTKVFDFHDEISQVPENNFVQPTHVTSQASGDTSLKPLVLSANSEP 1150

STRAIN=S288c / AB972;
MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaubin M.;
RT "Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome
RT VIII.";
RL SEQUENCE OF 265:2077-2082 (1994).
RN [3]
RP STRAIN=S288c;
RC STRAIN=S288c;
RX MEDLINE=88111539; PubMed=3322809;
RA Watts F.Z., Sheils G., Orr E.;
RT "The yeast MYO1 gene encoding a myosin-like protein required for cell
RT division.";
RL EMBO J. 6:3499-3505 (1987).
RT
CC -1- FUNCTION: Required for cell division.
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; X53947; CAA37894.1; -;
DR EMBL; X06187; CAA29550.1; -;
DR EMBL; U10399; AAB88872.1; -;
DR PIR; S46773; S46773.
DR HSSP; P08799; 1MND.
DR
DR GerMOnline; 139340; -;
DR SGD; S0001065; MYO1.
DR
DR GO; GO:0000142; C:contractile ring (sensu Saccharomycetes); IDA.
DR GO; GO:0007120; P:axial budding; IMP.
DR GO; GO:0000310; P:cytokinesis; IMP.
DR GO; GO:0006970; P:response to osmotic stress; IMP.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR
DR Myosin; Actin-binding; ATP-binding; Coiled coil.
FT DOMAIN 1 793 MYOSIN HEAD-LIKE.
FT DOMAIN 794 823
FT FT COILED COIL (POTENTIAL).
FT DOMAIN 856 1911
FT NP_BIND 180 187 ATP (BY SIMILARITY).
FT DOMAIN 460 529 ACTIN-BINDING (BY SIMILARITY).
FT FT CONFLICT 36 36 K -> I (IN REF. 1).
FT CONFLICT 46 46 I -> T (IN REF. 1 AND 3).
FT FT CONFLICT 59 59 V -> S (IN REF. 1 AND 3).
FT CONFLICT 86 86 L -> F (IN REF. 1).
FT FT CONFLICT 330 330 MISSING (IN REF. 1 AND 3).
FT FT CONFLICT 343 343 N -> S (IN REF. 1 AND 3).
FT CONFLICT 421 426 QAKFI -> TKLSSL (IN REF. 1).
FT CONFLICT 515 515 D -> S (IN REF. 1).
FT CONFLICT 529 535 SKGPPTG -> ARGHRD (IN REF. 1 AND 3).
FT FT CONFLICT 541 541 D -> V (IN REF. 1 AND 3).
FT FT CONFLICT 550 551 TD -> LM (IN REF. 1).
FT CONFLICT 573 573 R -> A (IN REF. 1).
FT CONFLICT 582 582 H -> D (IN REF. 1).
FT CONFLICT 588 599 EYTVGEWLSKNK -> NTLKWAGPKT (IN REF. 1).
FT CONFLICT 599 599 MISSING (IN REF. 3).
FT CONFLICT 627 632 EKSSSA -> GKNLLVC (IN REF. 1 AND 3).
FT FT

| | |
|----|--|
| DR | SMART; SM00242; MYSC; 1. |
| DR | PROSITE; PS50096; IQ; 1. |
| KW | Myosin; Actin-binding; ATP-binding; Coiled coil. |
| FT | DOMAIN 1 793 MYOSIN HEAD-LIKE. |
| FT | DOMAIN 794 823 IQ. |
| FT | DOMAIN 856 1911 COILED COIL (POTENTIAL). |
| FT | NP BIND 180 187 ATP (BY SIMILARITY). |
| FT | DOMAIN 460 529 ACTIN-BINDING (BY SIMILARITY). |
| FT | CONFLICT 36 36 K -> I (IN REF. 1). |
| FT | CONFLICT 46 46 I -> T (IN REF. 1 AND 3). |
| FT | CONFLICT 59 59 V -> S (IN REF. 1 AND 3). |
| FT | CONFLICT 86 86 L -> F (IN REF. 1). |
| FT | CONFLICT 330 330 MISSING (IN REF. 1 AND 3). |
| FT | CONFLICT 343 343 N -> S (IN REF. 1 AND 3). |
| FT | CONFLICT 421 426 QAKAFI -> TKLSSL (IN REF. 1). |
| FT | CONFLICT 515 515 D -> S (IN REF. 1). |
| FT | CONFLICT 529 535 SKGPPTG -> ARGHRD (IN REF. 1 AND 3). |
| FT | CONFLICT 541 541 D -> V (IN REF. 1 AND 3). |
| FT | CONFLICT 550 551 TD -> LM (IN REF. 1). |
| FT | CONFLICT 573 573 R -> A (IN REF. 1). |
| FT | CONFLICT 582 582 H -> D (IN REF. 1). |
| FT | CONFLICT 588 599 EYVEGEWSKNK -> NTLWKAGYPKT (IN REF. 1). |
| FT | CONFLICT 599 599 MISSING (IN REF. 3). |
| FT | CONFLICT 627 632 EKSSSA -> GKNLLVC (IN REF. 1 AND 3). |

RL Neuron 9:417-428 (1992).
 CC -!- TISSUE SPECIFICITY: GROWTH CONES OF EMBRYONIC VERTEBRATE NEURONS.
 CC -!- DEVELOPMENTAL STAGE: IS EXPRESSED IN THE NEURULA AND PERSIST
 CC DURING EMBRYOGENESIS IN THE BRAIN, CRANIAL NERVES, AND SPINAL
 CC CORD.
 CC -!- SIMILARITY: Belongs to the intermediate filament family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M99387; AAA49966.1; -;
 CC PIR; JH0720; JH0720.
 CC InterPro; IPR001664; IF.
 CC Pfam; PF00038; filament; 1.
 CC PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone.
 FT DOMAIN 1 12
 FT HEAD.
 FT DOMAIN 13 314
 FT ROD.
 FT DOMAIN 315 1744
 FT TAIL.
 FT DOMAIN 8 48
 FT COIL 1A.
 FT DOMAIN 49 60
 FT LINKER 1.
 FT DOMAIN 61 156
 FT COIL 1B.
 FT DOMAIN 157 179
 FT LINKER 12.
 FT DOMAIN 180 193
 FT COIL 2A.
 FT DOMAIN 194 199
 FT LINKER 2.
 FT DOMAIN 200 314
 FT COIL 2B.
 SQ SEQUENCE 1744 AA; 199561 MW; 6502EAC9F6C4E93 CRC64;
 Query Match 3.4%; Score 142; DB 1; Length 1744;
 Best Local Similarity 17.1%; Pred. No. 8.3;
 Matches 144; Conservative 145; Mismatches 308; Indels 244; Gaps 37;
 QY 8 QARTKNNRVSYDQKQATQKNTLTPDEVSKR-----EGINAE-QIVIKI---TDQ 56
 DB 659 KKKVPLENEYIPVSKDDLTETSHLENDSESQSFDSKLFENKSTEDQLITNLKNTQE 718
 QY 57 GYVTSHGHDHYHYNGKVPYDAI-----SELLMKDPN-----YKLKBDIVN 99
 DB 719 NIFQSNQEHLE---NLEFDSVVPDVKWYPOENLLSEENVYGDGLVQMATDENIIN 774
 QY 100 EVKGGVIVKGVYVYLKDAHADNVRTKEH-----NRQKQSHSQRHGGTFRNDGAVA 155
 DB 775 Q-----SSDQLLLSDHSHHEETKTSIATVEHNRMESEHAE-----VDKSE 816
 QY 156 LARSQRYTTDDGYIFNASDITETDGVIVPHGDHYHVPKPNLSASELAAAEFLSGR 215
 DB 817 IPVEISENVSEIHEISDVEDTKQAF----- 845
 QY 216 GNLSNRTYRRQNSDNTSTNVPVSNPPTNTNNTSNNSQASQNDISLLKQLY 275
 DB 846 ---EDERVEQINQNNQEST-----VDLDGSVYSQENSLQLEDEVSIEQIEKDFE--- 894
 QY 276 KLPLSQRHVESDGL--VFDPQAITSRTAGVAPHGDHYHFFPYQSOMSELEERARIPL 333
 DB 895 ---INEQELKSDQIRAEFTTEV-----DH-QVDFMQEQSFEREVGQLNNI 938
 QY 334 R-----YRSNHWVPSRPEQSPQTPPEPSFGPOPAPNLKIDGNSLSVQLVRKVGEGYVF 389
 DB 939 KQEVYDLYQNY-----DEDSQNNDEPO-----ELES-CDLQEQKIKLEENQLS 981
 QY 390 EKGISRY-----VPFAK-----DLPSETVKNL-----ESKLSQESVSHTL 425
 DB 982 ENEGNFGNGNDIEEFSQGYDTEICQETIGNQVSAQLLCESDINQDKLSMEDEEQNN 1041
 QY 426 TAKKENVAPR-----DQFYDKAYNLLTEAHKALFKXNGRNSDFQALD- 468
 DB 1042 PETEDNIGLEQESQDNTSRNEGTFKFSQECVVFVFPEDMSDKSEYSGQEDLDKQVTF 1101

QY 469 -----KLLER-----LNDESTNKEKLVDDL--AFLAPITHPERLCKPMS 506
 DB 1102 SLNEQANNDLLEKEEVILHHADDQRSVNDIITDEKLSERIIDNELATVDVNESLAANKE 1161
 QY 507 QLEYTEDEVRIAQL-----ADKYTTSDGYIFDEHDIISDEGDAYVTTHMGSHWIGKD 559
 DB 1162 QVDLFTDEYAVDNDVGMQDDSGQYQTKEDLFDVGNII-----EKIEIQOT 1208
 QY 560 SLSDKEKVAQAQYTKKGILPPSPDADVKANPTGDSAAAIYNRVKGEKRI-PLVRLPYMV 618
 DB 1209 SILNQE-ICERVNDVEDISGEAKNESVEMNDVDLVPEA--KVTGDEQISPLQDEKLN 1265
 QY 619 EHTVEVK--NGNLIIPHDKDHYHNKIPA-----WFDHITYKAPNGYTLDELFAIKYVVEHP 672
 DB 1266 ETMEDTKNDGQCLCEKENETEVIETVDSQPQATDLSHDAGRELTVDQNSANLQ-PCENP 1324
 QY 673 DERPHSNDGNGNASEHVLGKDHSDPKNKFKADEEPVEET---PAEPEVPQVETEKVEA 729
 DB 1325 TKT-----LIAHHI-----EYETVADSDLESTEEQVQETERIPPKPDSKMNENSES 1372
 QY 730 Q 730
 DB 1373 E 1373
 RESULT 23
 MOT3 YEAST STANDARD; PRT; 490 AA.
 ID MOT3 YEAST
 AC P54785;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein MOT3/HMS1.
 GN MOT3 OR HMS1 OR YMR070W OR YMR916.09.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Madison J., Winston F.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=97313268; PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT XIII.";
 RL Nature 387:90-93 (1997).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U25279; AAC49982.1; -;
 CC EMBL; Z48952; CA88795.1; -;
 CC PIR; S52830; S52830.
 CC GeneOnline; 142736; -;
 CC TRANSFAC; T03448; -;
 CC TRANSFAC; T03500; -;
 CC SGD; S0004674; MOT3.
 CC GO; GO:0005634; C:nucleus; IDA.

Db 212 RNLDQWRPSSLSQTSNTPHAANPSFPGTIVHTNFRPREGGCHRRSTGSL----SVG 268
QY 382 KVGEVYFEKGISR-YVFAKDLPSFTVKNLESKLSQESVSHTLFAKKENVAPRDQEFY 440
Db 269 SSGSGFSSGSGGNPKNLFSPYQSSIPAL---LAERRLVTGILIVSKKN---RSDAFV 322
QY 441 DRAYNLLTEAHKALFXNKGNN---SDFOALDKL-----LELNDESNKEK-----483
Db 323 SVDG---LDAEFTCGSKORNRALGDDVAIEILLDVEVWAGLKEEENRRKDPIDSTRG 379
QY 484 LVDDLAFAPITHPERLCKPNSQIETBYDEVRIAQLADKYTTSDGYIFDEHDIISDEGD 543
Db 380 SPFNLRIDAIVPEVPPQR-----SAIKARDEQVEGQTL-----FLLDQKLGADKXP 426
QY 544 AYVTPHMGSHHWGKDSLSDEKVAQYKKGILPPSPDADVK-----ANPTG 593
Db 427 KYA---GH-----VAVLQAPQGVFSGTGLILRPSSAANKERQTSNGGSSNNSG 475
QY 594 DSAATYNNRVKGEKRIPLVRLP 615
Db 476 NDKPKIWNFKPSDKRVFLTAIP 497

RESULT 21
D7_DICDI
ID D7 DICDI STANDARD; PRT; 850 AA.
AC P54682;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CAMP-inducible prespore protein D7 precursor.
GN D7.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=95080502; PubMed=7988791;
RA Agarwal A., Sloger M.S., Oyama M., Blumberg D.D.;
RT "Analysis of a novel cyclic Amp inducible prespore gene in
RT Dictyostelium discoideum: evidence for different patterns of cAMP
RT regulation."
RL Differentiation 57,151-162(1994).
CC -!- DEVELOPMENTAL STAGE: Expressed specifically in the prespore cells.
CC -!- INDUCTION: By CAMP.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U25143; AAA73514.1; --
DR DictyBase; DDB0001927; D7.
KW Sporulation; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 850 CAMP-INDUCIBLE PRESPORE PROTEIN D7.
FT DOMAIN 470 475 POLY-GLN.
FT DOMAIN 555 568 POLY-ASN.
FT DOMAIN 728 738 POLY-GLN.
SQ SEQUENCE 850 AA; 95343 MW; 13BA634CCE7AA502 CRC64;

Query Match 3.4%; Score 142; DB 1; Length 850;
Best Local Similarity 16.7%; Pred. No. 3.1;
Matches 140; Conservative 122; Mismatches 296; Indels 280; Gaps 29;
QY 15 NNEVSVIDGKQATQK-TENL--TPDEVSKREGNAEQIV--IKITQGYVTSHGDIHHY 69
Db 24 SNQISDEVGKAINQKLNENIEKVEDVVQFENV-ANQVIEELKVERHQ-----RQEL 73

QY 70 NGKVPYDAIISBE-----LLMKDPNYKLKDEDIYNEKGVYKVDGKYVYVLKDAHA 123
Db 74 LGEVHDSLRDSATNYIWGLLDKIOSYLPKDNKVKSEAFSSGQNNNIGSIGSDTGA 133
QY 124 DNVRTKEEN--RQKEHSQHREGGTPRNDGVALARSQGRVTTDDGYTFNADIEDTG 181
Db 134 STSPQFQINGLSGASQSSGSGTGGTSDSK-----TTNEAIFSSKVSITTDQ 183
QY 182 DAYIVPHGDHYHYIYKNELSASELAFAFLSGRNLNSRTRYRRONSNTRTNWPVS 241
Db 184 ESII-----GVAITAKDSLGAITITGLGVSSYAKVGGQITNGRAQGVITG 231
QY 242 SNPGTIN-----TNT-----SNNSTNSQASOSNDIDSLKOLYKLPLSQRH- 283
Db 232 DNTGTVGRGAVTTAGAVANTVGEFLGGSRGTGSSSAGTVGNVLSDSYTSIGKIASNGNS 291
QY 284 -VESDGLVEDPAQITSRARGVAVPHGDH-----YHPIYPS 318
Db 292 LSETIGTGTTLAHTFAGTDSVGT-GPHIITKTFNLIAAGKFNSSQOYIDKSYGSIQS 350
QY 319 QMSELEERTARIIPURYSNHWVDSRPPQSPQPTPEPSPQPPAPNLKIDSNSSLSVQ 378
Db 351 DNEEIKKRL-----QSAHQQLQEQSP-----371
QY 379 LVKRVGEGYVFEKGISRVYFAKDLPSFTVKNLESKLSQESVSHTLTAKKENVAPRDQ 438
Db 372 -----AIIQSMKSEDLKLLDDDEVIRNTLKEMOIQRENDQIQOQNG 412
QY 439 FYDKAYNL-LTBRAKALFXNKGNSDFQALDKLRLNDESNTEKELVDDLLAFAPITH 497
Db 413 --DKEQLIDLQNRPGLYKNQ--QDLK-----QEKRAHQQLI-----446
QY 498 PERLCKPNSQIETBYDEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGSHHW 557
Db 447 -----NYELNLQEDQEQYELLDQ-----LYDEQQ-----471
QY 558 KDSLSDEKVAQAQYTKKGILPPSPDADVKANPTGDSAAIYNNRVKGEKRIPLVRLPYM 617
Db 472 ---QOPQVSNKQQLQEQIINSPE-----IQQ 497
QY 618 VERTVEVKNGLIIPHKDHYHNIKFAWFDHTYKAPNGYTLDELDFATIKYVVEHPDERPH 677
Db 498 LNLHNG-----PFQDDYHNDQTEELKD-----DDYFNDDQQLNNQGFENNVEEPD 543
QY 678 SNDGKNASEHVLGKDHSEDNKFKADEBPVEETPAEPVEPQVETEKVEAQLKEAEVL 737
Db 544 LNDANDNFEQVNN 591
QY 738 LAKVTDSLSKANATETLAGLRNNLTQIMDNNSINAAEKLALLKLGSPSSVSKEKI 795
Db 592 LQEVED-----APERLYEIEHNSNL-----NKAVQEAEEIERQQNGNGSPAVNSHI 638

RESULT 22
TANA_XENLA
ID TANA_XENLA STANDARD; PRT; 1744 AA.
AC Q01550;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tanabin.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tadpole head;
RX MEDLINE=92398961; PubMed=1524825;
RA Hemmati-Brivanlou A., Mann R.W., Harland R.M.;
RT "A protein expressed in the growth cones of embryonic vertebrate
RT neurons defines a new class of intermediate filament protein.";

```
QY 235 TNWPSVS-----NPGTNTNTSNN 254
RA : : : : :
DB 336 EEMKSIENFFSEDKAHHLLOFNKEFEVRHDLFEKKLOKHFDVAKDTUNVGLRNTTVLS 395
RA : : : : :
QY 255 SNT-----NSQASQSDIDSLLKQLYKLPLSQRH-----VESDGLV-- 290
RA : : : : :
DB 396 SNTETMLKQYEDIKENLEQKMSKSSKDEMAKTINELSVTKGLMGVQBELLTSSGNIQT 455
RA : : : : :
QY 291 -----FDPAQITRTARGVA-VPHGDHYHFIPIYSQSELEERTARIIPLYR 336
RA : : : : :
DB 456 ALVSEMNTROELLDDA---SQTAKNYASLENLVKAYKAEIVQSNEYERIKHLESER-- 510
RA : : : : :
QY 337 SNHWVPDRPQSPQPTPEPSGPGAPNLKIDSNSLSVLQVVKVGEYVFEKGLSR 396
RA : : : : :
DB 511 -----STLSSQKQITISLG-----TK 527
RA : : : : :
QY 397 YVFADLPSETVKNLESK---LSKQESVSHLTAKENVAPRQDFYD--KAYNLLTBH 451
RA : : : : :
DB 528 EAQYEDL---VKLEAKNIEISQISGKEQSLSKQENLSNELKVKVQDQLEKLNLTIT 583
RA : : : : :
QY 452 KALFANK--GRNSFOAL-----DKLLRLNDESTNKEKLVDDLLAFAPITHPELGLKN 505
RA : : : : :
DB 584 KSNYENKISSQNEIVKALVSENDTLQRIQQLVEIKENEQKD-----HTTKLEAFQKN 637
RA : : : : :
QY 506 SQIETEDEVIRIAQLADKYTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKDSLSKE 565
RA : : : : :
DB 638 EQLQ--KLVNVEVQI-----KAHELEEQNRHLKNCLEKKTGVEESLSVK 683
RA : : : : :
QY 566 KVAQAQYTKBKGLPPSPADVKA-----NPTGDSAAIYNVRKGEKRIPIVRLPY 616
RA : : : : :
DB 684 TL-----KQVIVLSEKQDITAEKLELDQNLSELEVTNQLQKQVQSKR-----ELEQ 733
RA : : : : :
QY 617 MVEHTEVKNGLNLIIPKHQHNKFNKAFNFDHDTYKAPNGYILEDLFAIKYVVEPDRP 676
RA : : : : :
DB 734 KIKELEEIKNRNEPSPKGTQN-----FTKPSDPSKXNATNSLFPNNSAAIHSPMKK 788
RA : : : : :
QY 677 HNSDGNWGNASHVLGKHDSHPKNF-----KADBPVETAEPEVQVPEVEKYEA 729
RA : : : : :
DB 789 PKVD-----HISKSRINSSKETSKEFNDEFDLSSSNDLELTNPSP-----IQIKPVRG 837
RA : : : : :
QY 730 QLKEA-----EVLAKVYDSSLK 747
RA : : : : :
DB 838 KIKGSGNCMKPPISRSKKLLLVEDQSLK 867
RA : : : : :
RESULT 20
ID STS5 SCHPO STANDARD; PRT; 1066 AA.
AC 074454; 013452;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein sts5.
GN STS5 OR SPCC16C4.09.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RX MEDLINE=97041712; PubMed=8886983;
RA Toda T., Niwa H., Nemoto T., Dhut S., Eddison M., Matsusaka T.,
RA Yanagida M., Hirata D.;
RT "The fission yeast sts5+ gene is required for maintenance of growth
RT polarity and functionally interacts with protein kinase C and an
RT osmosensing MAP-kinase pathway.";
RL J. Cell Sci. 109:2331-2342(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
```


DR PIR: C64232; C64232.
 DR TIGR: MG292; --; 1.
 DR HAMAP: MF_00036; --; 1.
 DR InterPro: IPR002318; trna-synt_2c.
 DR InterPro: IPR006193; trna-synt_Ala.
 DR Pfam: PF01411; trna-synt_2c; 1.
 DR PRINTS: PR00980; TRNASYNTHALA.
 DR TIGRFAMs: TIGR00344; alas; 1.
 DR PROSITE: PS50860; AA_TRNA_LIGASE_II_ALA; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 SQ SEQUENCE 900 AA; 104301 MW; AA54520BFB3949A2 CRC64;

Query Match 3.4%; Score 143.5; DB 1; Length 900;
 Best Local Similarity 19.7%; Pred. No. 2.9; Mismatches 250; Indels 159; Gaps 28;
 Matches 127; Conservative 109

QY 273 QLYKPLSRQHVESDGLVDPDPAQITRTARG----VAVPHGDHYHPIPSQMSSELEERIA 328
 Db 199 EIWNIVFSQFNNDGNGNYTELAKNIDTGAIBRLVSVLQNS-----PTNFDTDIFLKL 253
 QY 329 RII----PURYSNHWVPSRPEQSPQPTPEP-----SPGQPPAPNL 367
 Db 254 KIEAFPCPKYDPSYFTF-----DPQVKYEQSYFRIIADHFKAITFTTISEGVLEGN- 306
 QY 368 KIDNSLSVQLVR-----KVGEGYVPEEKISRYVFAKD-----LPSETVK- 409
 Db 307 ---ERNVVRLLRLRALACKQLQLNLAFIEKIIDEIIASVYNYOHLKAKNETVAKQVVL 363
 QY 410 ---NLESKLSKQESVSHTLTAKENVAPRDQEPYDKAYNLLTEAHKALPKNKGRNSDFQA 466
 Db 364 KEINAFNKITDGLVFEKSVKNTLTPTQLNLETYGFVVEIIRELVNQKGLTIDMTV 423
 QY 467 LKLL-----LERLNDESTNKEKLVDDLLAFAPITHPERLQKPNQSI--EYTEDVRIA 518
 Db 424 FQQLMAKHSRSKQNNQNTINFEKQINLVNFKTKSTFFYHKNKINAKVIGLGFENLVPVK 483
 QY 519 QLADK-----YTTSDGVIFDEHDI--SDEGDAVVT-----PHMGHSHW--I 556
 Db 484 ELNQGSGYVFDQTVIATSGGRYDEGSCINHSNNNDQKISFGQVFGKPNQKHFFYFLV 543
 QY 557 GKDSLSDK-----EKVAAQAYTKE---KGILPPSPDADVK---ANPTGDSAAIYN 601
 Db 544 GSFKLNQDVTLSHDETWRKLANNSHLEHLLHAALQKEIDPLIKQSGAFKSAQAATIDFN 603
 QY 602 RVKGEKRIPLVLPYVVEHTVEKNG-----NLIIPKDYH--HNI-KFAW 644
 Db 604 LNRHLTRNELEKVENKIRSLIKQKISSKEIFTFDEGSKLNAIAYPEEYSQHEILRVIR 663
 QY 645 FDDHTYK-----APNGYITLEDLFAFIKYVVEHPDERPHSDGWNASBHVILGKKDHS 698
 Db 664 FGDYSVELCGGTHVANTASIEOCFTIDFYSL-----GAGRWRIEISS---NET 709
 QY 699 PKNKFKADEPVEETPAEPE--VPQVETEKVBAQLKEAE-----VLLAKVTDSLLKANA 750
 Db 710 INNYLKAENQKLTQLKSELEKVLSDISSIFKVELKQLQRLDKFTLPEKITQ---LRDA 766
 QY 751 TETLAGLRNLTLLQINDNNSIMAEA-----EKLALLKGSNPSSV 790
 Db 767 SDTLIAKLNIDINQLTKNYKVSOQALALSIRKQLLSLVDENKSYV 811

RESULT 19

ID ZIPL_YEAST
 AC P31111; STANDARD; PRT; 875 AA.
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Synaptonemal complex protein ZIPL1.
 GN ZIPL1 OR YDR285W OR D9819.9.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=BR1824-3B;
 RX MEDLINE=93161412; PubMed=7916652;
 RA Sym M., Engebrecht J.A., Roeder G.S.;
 RT "ZIPL1 is a synaptonemal complex protein required for meiotic
 chromosome synapsis";
 RL Cell 72:365-378(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
 RA Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Required for meiotic chromosome synapsis and cell cycle
 progression. May act as a molecular zipper to bring homologous
 chromosomes in close apposition. ZIPL1 may encode the transverse
 filaments of the synaptonemal complex.
 CC -!- SUBCELLULAR LOCATION: SYNAPSED MEIOTIC CHROMOSOMES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L06487; AAA35239.1; --
 DR EMBL; U51031; AAB64474.1; --
 DR PIR; S70115; S70115
 DR GeneOnline; 140777; --
 DR SGD; S0002893; ZIPL1.
 DR GO; GO:0000795; C:synaptonemal complex; IDA.
 DR GO; GO:0007126; P:meiosis; IMP.
 DR GO; GO:0007129; P:synapsis; IMP.
 KW Nuclear protein; Meiosis; Coiled coil.
 FT DOMAIN 177 333 COILED COIL (POTENTIAL).
 FT DOMAIN 397 438 COILED COIL (POTENTIAL).
 FT DOMAIN 456 752 COILED COIL (POTENTIAL).
 FT CONFLICT 55 55 T -> A (IN REF. 1).
 SQ SEQUENCE 875 AA; 100035 MW; 674F12625CD9DDED CRC64;

Query Match 3.4%; Score 142.5; DB 1; Length 875;
 Best Local Similarity 18.9%; Pred. No. 3;
 Matches 176; Conservative 117; Mismatches 304; Indels 333; Gaps 41;

QY 3 ELGLQART-VKENRVSYIDGKQATKTENITPDVEKRGINAIEQIVIKITDQGYVTS 61
 Db 86 EIGSPKTTSTDOYNRL----KNDVALENDTDEDTEITVEVSEGVAKETKE----S 137
 QY 62 HGD-----HYHNYNGKVPYDAII-----SEELLM----- 85
 Db 138 HGDPNDSSETTLKDSKMHYEYTMGTNGKAPLHTSINNSSTSSNDVLLFAFTTQRTICSNLKE 197
 QY 86 ----KDPNKKK-----DEDIVNEVKGVIKVDGKYVYLKD-----AAHADN 125
 Db 198 LQKQODNAKLKVRLOSVASNSDKINE-----KV-GKYSCLTQLERATLTSHKNN 249
 QY 126 VRTKEEINRQKQESHQSHREGGTPRNDGAVALARSQGRYTTDDGYTFNADSIIDETG---- 181
 Db 250 QETKLKDLRQNHLYQRRISG-----PKTS---TENLKTINDLGKNKK 290
 QY 182 ---DAYIVPHGDHYYIPKNELSGASELAAAEAFSLGRGNLSNRTYRRQNSD-----NTSR 234
 Db 291 EADAELMKKGKIEYL-KRELDCCS-----GQLSEKI---KNSSLIQEMGKNR 335

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata: Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1A), FUNCTION, PHOSPHORYLATION, AND
 RP INTERACTION WITH MAPK8; MAPK9; MAPK10; MAP2K4 AND MAP3K1.
 RC TISSUE=Brain;
 RX MEDLINE=99455010; PubMed=10523642;
 RA Ito M., Yoshioka K., Akechi M., Yamashita S., Takamatsu N.,
 RA Sugiyama K., Hibi M., Nakabeppu Y., Shiba T., Yamamoto K.-I.;
 RA "JSAPl, a novel jun N-terminal protein kinase (JNK)-binding protein
 RT that functions as a scaffold factor in the JNK signaling pathway";
 RT Mol. Cell. Biol. 19:7539-7548 (1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1A; 1B; 1C; 1D AND 1E), AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=20480689; PubMed=11024282;
 RA Ito M., Akechi M., Hirose R., Ichimura M., Takamatsu N., Xu P.,
 RA Nakabeppu Y., Tadayoshi S., Yamamoto K.-I., Yoshioka K.;
 RA "Isoforms of JSAP1 scaffold protein generated through alternative
 RT splicing";
 RT Gene 255:229-234 (2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1C AND 3A), FUNCTION, INDUCTION, TISSUE
 RP SPECIFICITY, SUBCELLULAR LOCATION, PHOSPHORYLATION, MUTAGENESIS OF
 RP ARG-205; PRO-206; THR-207; SER-208; LEU-209; THR-266; THR-276 AND
 RP THR-287, AND INTERACTION WITH MAPK8IP2; MAPK8; MAPK9; MAPK10; MAP2K7
 RP AND MAP3K11.
 RC STRAIN=C57BL/6; TISSUE=Brain, and Heart;
 RX MEDLINE=20094982; PubMed=10629060;
 RA Kelkar N., Gupta S., Dickens M., Davis R.J.;
 RA "Interaction of a mitogen-activated protein kinase signaling module
 RT with the neuronal protein JIP3";
 RT Mol. Cell. Biol. 20:1030-1043 (2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1C), SUBCELLULAR LOCATION, AND INTERACTION
 RP WITH KLC1.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=20560743; PubMed=1106729;
 RA Bowman A.B., Kamal A., Ritchings B.W., Philp A.V., McGrail M.,
 RA Gindhart J.G., Goldstein L.S.B.;
 RA "Kinesin-dependent axonal transport is mediated by the Sunday Driver
 RT (SYD) protein";
 RT Cell 103:583-594 (2000).
 RN [5]
 RP SEQUENCE OF 1240-1337 FROM N.A.
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.-I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Murny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [6]
 RP INTERACTION WITH KLC.
 RC TISSUE=Brain;
 RX MEDLINE=21135887; PubMed=11238452;
 RA Verhey K.J., Meyer D., Deehan R., Blenis J., Schnapp B.J.,

RA Rapoport T.A., Margolis B.;
 RT "Cargo of kinesin identified as JIP scaffolding proteins and
 RT associated signaling molecules";
 RL J. Cell Biol. 152:959-970 (2001).
 CC -1- FUNCTION: The JNK-interacting protein (JIP) group of scaffold
 CC proteins selectively mediates JNK signaling by aggregating
 CC specific components of the MAPK cascade to form a functional JNK
 CC signaling module. May function as a regulator of vesicle
 CC transport, through interactions with the JNK-signaling components
 CC and motor proteins.
 CC -1- SUBUNIT: Forms homo- or heterooligomeric complexes. The central
 CC region of Mapk8ip3 interacts with the C-terminal of Mapk8ip2 but
 CC not Mapk8ip1. Binds specific components of the JNK signaling
 CC pathway namely Mapk8, Mapk9 and Mapk10 to the N-terminal region,
 CC Map2k4 and Map2k7 to the central region and Map3k11 to the C-
 CC terminal region. Binds the TPR motif-containing C-terminal of
 CC kinesin light chain, pre-assembled Mapk8ip1 scaffolding complexes
 CC are then transported as a cargo of kinesin, to the required
 CC subcellular location.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; localised in the soma and
 CC growth cones of differentiated neurites and the Golgi and vesicles
 CC of the early secretory compartment of epithelial cells.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=6;
 CC Name=1c; Synonyms=3b;
 CC IsoId=Q9ESN9-1; Sequence=Displayed;
 CC Name=1a;
 CC IsoId=Q9ESN9-2; Sequence=VSP_002775, VSP_002777;
 CC Name=1b;
 CC IsoId=Q9ESN9-3; Sequence=VSP_002776, VSP_002777;
 CC Name=1d;
 CC IsoId=Q9ESN9-4; Sequence=VSP_002775;
 CC Name=3a;
 CC IsoId=Q9ESN9-5; Sequence=VSP_002778, VSP_002779;
 CC Name=1e;
 CC IsoId=Q9ESN9-6; Sequence=VSP_002776;
 CC -1- TISSUE SPECIFICITY: Highly expressed throughout many regions of
 CC the brain and at lower levels in the heart, liver, lung, testes
 CC and kidney. All isoforms have been identified in the brain,
 CC Mapk8ip3A is also expressed in the spleen and lung.
 CC -1- INDUCTION: Expressed in neurites 5 days following initiation of
 CC nerve growth factor Ngf induced differentiation. Ngf withdrawal
 CC results in the down-regulation of Mapk8ip3 protein by caspase-
 CC mediated cleavage.
 CC -1- SIMILARITY: Belongs to the JIP scaffold family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB005662; BAB58974.1; -
 CC EMBL; AB043124; BAB16675.1; -
 CC EMBL; AB043125; BAB16676.1; -
 CC EMBL; AB043123; BAB16674.1; -
 CC EMBL; AB043129; BAB16685.1; -
 CC EMBL; AB043126; BAB16685.1; JOINED.
 CC EMBL; AB043127; BAB16685.1; JOINED.
 CC EMBL; AB043128; BAB16685.1; JOINED.
 CC EMBL; AF178637; AAF26843.1; -
 CC EMBL; AF178636; AAF26842.1; -
 CC EMBL; AF262046; AAG36931.1; ALT_INIT.
 CC EMBL; BC004003; AAH04003.1; -
 CC MGD; MGI:1353598; Mapk8ip3.
 CC GO; GO:0005737; Cytoplasm; IEP.
 CC GO; GO:0019894; F:kinesin binding; IPI.
 CC GO; GO:0005078; F:MAP-kinase scaffold activity; IPI.
 CC GO; GO:0019901; F:protein kinase binding; IPI.
 CC GO; GO:0046328; P:regulation of JNK cascade; IDA.
 CC GO; GO:0016192; P:vesicle-mediated transport; IDA.

